NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

GEO-DEMOGRAPHIC ANALYSIS IN SUPPORT OF THE UNITED STATES ARMY RESERVE (USAR) UNIT POSITIONING AND QUALITY ASSESSMENT MODEL (UPQUAM)

by

Martin Lynn Fair

June 2004

Thesis Advisor: David H. Olwell
Second Reader: Samuel E. Buttrey

Approved for public release; distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2004	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE: Geo-Demographic Analysis in Support of the United States Army Reserve (USAR) Unit Positioning and QUality Assessment Model (UPQUAM)	**5. FUNDING NUMBERS**
6. AUTHOR MARTIN LYNN FAIR	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution is unlimited	**12b. DISTRIBUTION CODE** A

13. ABSTRACT *(maximum 200 words)*

Manning United States Army Reserve (USAR) units are fundamentally different from manning Regular Army (RA) units. A soldier assigned to a USAR unit must live within 75 miles or 90 minutes commute of his Reserve Center (RC). This makes reserve unit positioning a key factor in the ability to recruit to fill the unit. This thesis automates, documents, reconciles, and assembles data on over 30,000 ZIP Codes, over 800 RCs, and over 260 Military Occupational Specialties (MOSs), drawing on and integrating over a dozen disparate databases. This effort produces a single data file with demographic, vocational, and economic data on every ZIP Code in America, along with the six year results of its RA, USAR, sister service recruit production, and MOS suitability for each of the 264 MOSs. Preliminary model development accounts for about 70% recruit production variation by ZIP Code. This thesis also develops models for the top five MOSs to predict the maximum number of recruits obtained from a ZIP Code for that MOS. Examples illustrate that ZIP Codes vary in their ability to provide recruits with sufficient aptitude for technical fields. Two subsequent theses will use those results. One completes the MOS models. The second uses the models as constraints in an optimization model to position RCs. An initial version of the optimization model is developed in this thesis. Together, the three theses will provide a powerful tool for analysis of a strategic-based optimal reserve force stationing.

14. SUBJECT TERMS Analysis, Regression, Statistics, Unit Stationing, Recruiting, Marketing, Data Manipulation, Data Mining, Database Management, and Clementine Software	**15. NUMBER OF PAGES** 144
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited.

GEO-DEMOGRAPHIC ANALYSIS IN SUPPORT OF THE UNITED STATES ARMY RESERVE (USAR) UNIT POSITIONING AND QUALITY ASSESSMENT MODEL (UPQUAM)

Martin Lynn Fair
Lieutenant Colonel, United States Army
B.S., Indiana University of Pennsylvania, 1983
M.S., Indiana University of Pennsylvania, 1994
M.S.I.E., University of Louisville, 1998

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2004

Author: Martin Lynn Fair

Approved by: David H. Olwell
Thesis Advisor

Samuel E. Buttrey
Second Reader/Co-Advisor

James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Manning United States Army Reserve (USAR) units is fundamentally different from manning Regular Army (RA) units. A soldier assigned to a USAR unit must live within 75 miles or 90 minutes commute of his Reserve Center (RC). This makes reserve unit positioning a key factor in the ability to recruit to fill the unit.

This thesis automates, documents, reconciles, and assembles data on over 30,000 ZIP Codes, over 800 RCs, and over 260 Military Occupational Specialties (MOSs), drawing on and integrating over a dozen disparate databases. This effort produces a single data file with demographic, vocational, and economic data on every ZIP Code in America, along with the six year results of its RA, USAR, sister service recruit production, and MOS suitability for each of the 264 MOSs.

Preliminary model development accounts for about 70% recruit production variation by ZIP Code. This thesis also develops models for the top five MOSs to predict the maximum number of recruits obtained from a ZIP Code for that MOS. Examples illustrate that ZIP Codes vary in their ability to provide recruits with sufficient aptitude for technical fields.

Two subsequent theses will use those results. One completes the MOS models. The second uses the models as constraints in an optimization model to position RCs. An initial version of the optimization model is developed in this thesis.

Together, the three theses will provide a powerful tool for analysis of a strategic-based optimal reserve force stationing.

THIS PAGE INTENTIONALLY LEFT BLANK

DISCLAIMER

The views expressed in this document are those of the author. They do not reflect official policy, regulatory requirement, position of Department of Defense, or the government. The reader should understand the author decided discrepancies of data elements. While every effort was taken to ensure error-free computer programs, logic, and computational errors, the reader must use the information with discretion. Any usage, beyond the scope of this thesis, is done at risk of the user, and validation should be made prior to future application. The terms he, him, or himself refers to both genders masculine and feminine.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF EQUATIONS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS

ACRONYM	DEFINITION
AFQT	Armed Forces Qualification Test
AIT	Advanced Individual Training
ANOVA	Analysis of Variance
ARCOM	Army Reserve Command
ARNG	Army National Guard
ASG	Area Support Group
ASVAB	Armed Forces Vocational Aptitude Battery
BLS	Bureau of Labor and Statistics
BRAC	Base Realignment And Closure
BT or BCT	Basic Training or Basic Combat Training
CAR	Chief of the Army Reserve
CMA-R	Competitive Market Analysis – Reserve
CONUS	Continental United States
CONUSA	Continental United States Army
CPS	Current Population Survey
FEMA	Federal Emergency Management Area
FIP	Federal Information Partnership
FY	Fiscal Year
GP	General Population
LAU	Local Area Unemployment
LM	Linear Model
LP	Linear Programming
LSCAT	Line Score Category
MA	Military Available
MEPS	Military Entrance and Examination Processing Station
MOS	Military Occupational Specialty

ACRONYM	DEFINITION
MSS	Market Supportability Study
NLP	Non-Linear Program
NMA	National Market Analysis
NPS	Non-Prior Service
OCAR	Office of the Chief of the Army Reserve
OMAR	Operations and Maintenance Army Reserve
POM	President’s Objective Memorandum
PS	Prior Service
RA	Regular Army
RC	Reserve Center
RSC	Regional Support Command
RDBMS	Relational DataBase Management System
TPU	Troop Program Unit
UIC	Unit Identification Code
UPQUAM	Unit Positioning and QUality Assessment Model
USAR	United States Army Reserve
USARC	United States Army Reserve Command
USAREC	United States Army Recruiting Command
USBC	United States Bureau of the Census
USPS	United States Postal Service
ZIP	Zone Improvement Plan

ACKNOWLEDGMENTS

To my family, my sincere thanks and apologies for missed times and opportunities. Thank you for your understanding, patience, and support while I attended to my educational goal and professional development. To my wife, Susan, I would not have been successful without you in my life. A very loving thank you. You complete me. Thank you for our children, your patience, inspiration, time, encouragement, friendship, and understanding. May the future bring us continued happiness, fruitfulness, understanding, and a love that grows as our relationship. One that endures all of our endeavors as time endures us ... I love you!

To My Mason, My Morgan, and My Marlee, I am so proud of you still being excited when Daddy comes walking through the door even when you may have not seen him all day or long into the next day. You all have certainly blossomed into very good children. I am certain you all will become more successful than I in your future goals. Please continue to strive to be your best in all that you endeavor. Your education is the key to success. Always set attainable goals for yourself and be tenacious at completing them. You will forever have your parents' support and love.

I would like to thank the following individuals for without their existence, support, or input; this thesis would have not been possible: To my Mother, Janice V. Fair; deceased Father, Floyd A Fair; and Adopted Step-Father, Gary A. Henry; for my birth, rearing, values, and principles.

Mom, thank you for teaching me the difference between right and wrong, good and bad, and the wisdom to know the difference. We have come a long way in such a short time. Thank you for letting me be me and making the mistakes necessary to develop my character. I hope I have made you proud. Mom, I am forever grateful for your patience, encouragement, understanding, respect, kindness, and uncanny ability to get anything and everything done without failure. I cannot express or say everything I want to about you...I just want you to know I love you Mom!

Dad, thank you for demonstrating there are many ways to show love even though I did not understand them at the time. My only regret is that I wish we could have had more time to learn from each other. I wish you could be here to see your grandchildren…they would make you proud! I'm sorry, I did tell Mom about my graduation gift! I hope I have made you proud. You will be forever missed!

Gary, thank you for ALWAYS being there. You have an uncanny ability to brighten anybody's day. I only wish I could be as patient as you are. You demonstrate and are the epitome that love has no bounds. I am glad you entered my life and always gave unselfishly of yourself. I have certainly learned many valuable lessons from you and your example. I never looked for a replacement for a Father, but my choice would be you. I hope I have made you proud.

To my Brother and Sister, Ricky and Lisa; for always being there when we needed each other. Thank you both for putting up with my over zealousness at times. I hope I have made you both proud. We have certainly built each other's character over the years. My choice of siblings would certainly be the both of you. Without you both, this thesis would not be possible. I love both of you!

To my Grandmother, Helen M. Clawson, for my rearing, values, and principles. Thank you for teaching me the difference between right and wrong, good and bad, and the wisdom to know the difference. I hope I have made you proud. Grandma, I am also forever grateful for your patience, encouragement, understanding, and always being there when needed. I Love You Grandma!

To all my family, soldiers, co-workers, bosses, and friends; thank you for developing my character and giving me the opportunity to be a part of your life as you have been a part of mine.

To all the faculty, staff, and administration at the Naval Postgraduate School; past and present; I came seeking knowledge, you not only gave me that knowledge, but gave of yourselves. I hope I represent the knowledge, character, and professionalism this institution exemplifies. Thank you for the knowledge and experience. I am forever grateful.

A special thanks to all my teachers, instructors, professors; past, present, and future; without your insight, knowledge, patience, and understanding this thesis would not be possible.

A very special thanks to Mr. Rodderick Lunger and Major Michael Kamei, Program, Analysis, and Evaluations Directorate, United States Army Recruiting Command, Fort Knox, Kentucky; Mrs. Susan Fair, Field Representative, United States Census Bureau, Mrs. June Grillo, Senior Field Representative, United States Census Bureau, and Mr. Jamey Christy, Regional Director, United States Census Bureau; and Major Ward Litzenburg, Office of the Chief of Army Reserve for all the data support and the professional persevering manner in which you provided the data. Without your dedication and assistance this thesis would not be possible. Thank you.

A very special thanks to Dr. Charles R. Maderer; my professor, advisor, friend, and mentor. Without your patience, encouragement, friendship, knowledge, and understanding this thesis would not be possible. You truly are one of a kind Chuck. You are the best --- Ah - right ... I love you my friend!

A very vey special thanks to the members of my thesis committee; Dr. David H. Olwell, and Dr. Samuel E. Buttrey. Dave and Sam, thank you for your patience, time, encouragement, friendship, knowledge, input, and understanding. Thank you for your dedication, professionalism, and positive fervor of teaching and facilitating. I would have never completed this thesis without your tenacity. I am forever grateful. Because of you, I am better.

There is an Old Irish Blessing I love to hear and use. I dedicate the following prose to all students, all faculty, soldiers (friend or foe), and all mankind of which I have met or will meet during my lifetime -- All of which I consider my friends. It goes something like this; "May the rain fall softly upon your shoulders, may the sun always shine upon your face, may the wind always be at your back, and until we meet again, may the Lord lift you and keep you in the palm of His hand."

May God Always be at the Helm!
Take Care, God Bless, and God's Speed Always!

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Trained and ready units are the key to the success of America's Armed Forces. The drawdown of United States Armed Forces over the past decade and a half causes great reliance on the Reserve Components. With this increased reliance, unit fill becomes increasingly important to unit deployment schedules and Homeland Security. Unfilled units degrade personnel and training readiness. This thesis develops a three-phase modeling process that will greatly assist with the analysis of this readiness issue.

Manning United States Army Reserve (USAR) units is fundamentally different than manning Regular Army (RA) units. A soldier assigned to a USAR unit must live within 75 miles or 90 minutes of his Reserve Center (RC). This makes USAR unit positioning a key factor in the ability to recruit to fill the unit.

This model addresses this problem by looking at specific demographic, vocational, and other ZIP Code factors of interest. This thesis is Phase I of a three theses effort to address this problem. These three phases are:

Phase I: Process Definition, Data Collection, and Data Scrubbing.

Phase II: MOS Build – Populate Data Fields for the Optimization Model.

Phase III: Construct and Complete the Optimization Model.

Since the entire model is a huge undertaking, the focus of this thesis is Phase I. Prior to an analysis, data collection and data scrubbing take an enormous amount of time and effort. In this thesis, we assemble the data on over 30,000 ZIP Codes, over 800 RCs, and over 260 Military Occupational Specialties (MOSs), drawing on and integrating over a dozen disparate data bases. Phase I is an exercise in data mining, data manipulation, data acquisition, and data sourcing identification.

This effort produced a single table with demographic, vocational, and economic data on every ZIP Code in America, along with the six-year results of RA, USAR, and Sister Service recruit production. Data was also obtained on the quality of each recruit and his suitability for each of the 264 Army MOSs.

Preliminary modeling developed a model that accounts for about 70% of the variation in recruit production by ZIP Code. Models for the top five USAR MOSs were also developed to predict the maximum number of recruits obtained from a ZIP Code for that MOS. ZIP Codes vary in their ability to provide recruits with sufficient aptitude for technical fields, and this is illustrated in this thesis with examples. This modeling gives new explanatory and predictive capability. Surprisingly, unemployment rates had a small inverse effect on these five models. The unemployment rate is statistically significant, but may not be practically significant.

The second thesis in the series will develop models for all 264 MOSs and analyze them for commonalities and differences that reveal insights about recruit production for the USAR. This will also identify the regional propensity of the market to join the USAR. The third thesis will use those models as constraints in a mixed integer linear program that positions the RCs to maximize their ability to man their units. The assignment of RC market ZIP Codes to maximize unit fill rates leads to increased unit readiness. This thesis creates an initial version of this program.

This thesis automates the process of assembling and reconciling key data files using a commercial data-mining package called Clementine. We document that process so that future analysts can avoid the near three man-months of work to create an updated master data file with its over 30,000 by 430 cells. This is a major contribution.

These results support the solution of the unit fill rate problem and address many of the issues associated with determining the appropriate demographic, economic, and vocational factors of RC markets. Together these three theses will provide a powerful tool for analysis of optimal reserve force stationing. This will greatly improve the readiness of the Reserve Components, unit deployment schedules, and Homeland Security.

I. INTRODUCTION, BACKGROUND, AND SOURCE

A. INTRODUCTION

Trained and ready units have been the key to success of America's Armed Forces. Without a trained and ready force, we cannot support and defend the nation. The Army Accession Command has the responsibility to fill the ranks of the Army. One of its subordinate units is the United States Army Recruiting Command (USAREC). USAREC has the responsibility to achieve both its Regular Army (RA) and United States Army Reserve (USAR) annual accession missions. Without soldiers, the armed forces cannot begin to be ready and trained. Unit fill is the first step in achieving ready, trained, and deployable units.

This thesis focuses on recruitment quality and unit placement, with respect to the population, to meet force structure objectives. This thesis develops a model to analyze the complex process of filling the USAR Troop Program Unit (TPU) vacancies. The model determines factors associated with unit fill rates by Military Occupational Specialty (MOS). The model looks at unit positioning, assesses the quality of potential recruits, and includes demographic considerations to determine potential success in the market by MOS. The MOS fill rate is as follows:

$$FILL\_RATE_{MOS\&SKILL\_LEVEL} = \frac{ON\_HAND_{MOS\&SKILL\_LEVEL}}{AUTHORIZED_{MOS\&SKILL\_LEVEL}}$$

Equation 1.1: MOS Fill Rate Equation

For example, if we have a unit with 15 63B10 (skill level 1) authorizations and 5 63B20 (skill level 2) authorizations and it had 10 63B10s on-hand and 3 63B20s on-hand, the fill rate for skill level 1 and 2 63Bs would be 0.667 and 0.600, respectively. Modeling the process of filling unit vacancies will greatly assist in accessing the requisite number of young men and women soldiers for America's Army.

I call this model the Unit Positioning and QUality Assessment Model (UPQUAM). UPQUAM is a marketing and enlisted quality assessment tool used for conducting strategic USAR unit positioning and quality assessment for USAREC and the United States Army Reserve Command (USARC).

The USARC consists of 10 Regional Support Commands (RSCs) comprising over 4200 individual TPUs plus 4 other Army Commands (ARCOMs) to support its mission responsibilities. For National Security and Homeland Defense, these RSCs are aligned with the Federal Emergency Management Areas (FEMAs).

The location of US Armed Forces Reserve units plays an important role in Homeland Security issues as well as National defense posturing for success on the battlefield. Note that the former Continental United States Armies (CONUSAs) have been realigned with the FEMAs. The reason was to provide a support infrastructure for Homeland Defense in each FEMA. This analysis examines the relation between unit location and recruiting success. We desire to consider how to maximize the fill rate of USAR units through regression and optimization.

The model takes as inputs USAR unit structure, location, and historical quality of enlistment contracts. It uses a threshold value, for each MOS, based on Armed Forces Scoring Vocational Aptitude Battery (ASVAB) Line Score Categories (LSCATs). There are ten LSCATs which determine the minimum requirements for obtaining or qualifying for a particular MOS. The average LSCATs for each ZIP Code and Reserve Center (RC) will determine the type(s) of MOSs supported by the population surrounding the RC.

This thesis models the number of recruits a ZIP Code should produce, and the maximum number of recruits with sufficient skills for each MOS. This is a necessary input to the UPQUAM model, which will be completed in a subsequent thesis. The combined analysis will give insight as to the proper districting of RC areas, a specific location for USAR units throughout the US. The analysis illustrates the issues associated with unit vacancy fill problem of TPUs in the USAR.

B. BACKGROUND

One of many missions of the USAR is to recruit to fill its ranks. USAREC administers this responsibility by recruiting, assessing, and accessioning to fill USAR TPUs. Maintenance of quality soldiers for the USAR is a TPU responsibility. The recruiters' mission greatly hinges on the ability of the market (the population) to support the USAR units in their respective locations.

Filling RA and USAR units requires different approaches. Recruits filling RA units are accessed, attend training, and then are sent to their units worldwide without respect to their place of entry. USAR units are, normally, filled by personnel recruited within 75 miles or 90 minutes commuting time. This constraint is imposed to reduce the financial burden on soldiers. This geographical limitation, at times, may hamper unit fill. This occurs because personnel necessary to fill the unit are taken from a geographical region and there may or may not be sufficient numbers of qualified personnel in the region suited to join the units.

This analysis focuses on USAR force structure and the geographical constraints placed on units with respect to the local population. Filling unit vacancies comes at a price. Historically, fill rates of units (the percentage of required personnel in certain geographical locations) have not been at appropriate readiness levels.

There are two sets of qualified applicants, Prior Service (PS) and Non-Prior Service (NPS) personnel. These two pools of personnel form the available population. The Army considers the Military Available (MA) population those individuals aged 17-29.5 who are mentally, morally, and medically qualified for military service. The NPS set is those individuals aged 17-21 and the PS set is those individuals aged 22-29.5.

The USAR is ultimately responsible for filling its ranks. However, USAREC is responsible for recruiting the NPS set and the USAR is responsible for the PS set. PS personnel, as the name indicates, have previously served. To administer the PS responsibility, the USAR maintains a database of qualified soldiers to deploy when needed. The motivation for PS personnel to stay is greatly influenced by their respective

unit experiences. Since this is a TPU responsibility and not a focus of this study, we will not consider the PS set.

Instead, the analysis focuses on the NPS set. This is the harder set for data assembly and analysis. Recruitment for a particular position is based on its vacancy. Readiness, as previously stated, is a function of personnel. To have ready and trained units, the USAR must first train the personnel it recruits to perform specific tasks or missions. Recruits must have sufficient aptitude to be task trained, and are tested to see if they do.

The collection of skills for a position has an associated MOS. Soldiers receive MOS training in two phases. The first phase, indoctrination, is called Basic Training (BT). This is where soldiers receive training in basic combat skills. The second phase, the skill set for an MOS, is called Advanced Individual Training (AIT). Each position in a unit has an associated MOS and experience levels. Not all vacant positions in a unit are at a novice level. As a soldier gains experience and expertise, he becomes responsible for additional skills within his MOS.

The unique challenge for the USAR is the traveling constraint for unit personnel reporting for duty. As previously stated, this limit is currently 75 miles or 1.5 hours commuting time to the unit. Commuting distance for a unit headquartered in rural areas differs from those in suburban areas because of traffic. It may take just as much time to travel 25 miles in suburban areas as it does to travel 75 miles in rural areas. Therefore, geographical location of units with respect to the population is a major consideration.

Personnel with different skills may be more apt to join units demanding these skills. The Bureau of Labor & Statistics (BLS) and the United States Bureau of the Census (USBC) collects data about vocational aptitudes. This thesis considers eleven different vocational categories for the workforce. There is a clustering of USAR MOSs to these eleven vocational categories. We determine the inclusion of these vocational categories as we conduct a regression analysis.

Currently, the types and markets of some units do not align. Some local markets cannot adequately support the unit requirements. This is cause for concern, especially if the unit has a high priority for deployment. Unit fill is essential for readiness. Improving

the unit location with respect to the local market may make more effective use of the MA population. TPU alignment within its respective market should be such that the recruiting mission is attainable. TPU structure positioned to draw on the local vocations is one way to accomplish the recruiting mission. An extension of our model allows an optimal RC unit-stationing plan, and I discuss this in Chapter V.

The primary purpose of this thesis is to determine which demographic factors affect unit fill rates. There may be several causes for the lack of unit fill over time such as unit attrition, unit climate, market, recruitment efforts, population demographics, unemployment rates, quality, mission goals, or other factors. It is the responsibility of both USAREC and the USARC to determine what they individually and jointly can do about the lack of fill. These unit fill rates are key inputs into the larger position problem.

Insufficient unit fill itself gives no indication as to specific causes. If unit shortages are left unattended, the results can be devastating to Homeland and National Security. Currently policy and regulatory requirements incorporate some methods to relocate and reposition structure. There is a need for additional methods and policy to ensure unit fill. If this analysis proves beneficial, the Chief, Army Reserve (CAR) and USARC Force Structure personnel should adopt a strategy of repositioning structure in accordance with this analysis.

C. PROBLEM AND SOURCE

1. Underlying Problem

The complexity of this problem is too vast for one thesis. To manage the process, I will break it down into three components.

1. Phase I: Process & Model Definition, Data Collection, and Data Scrubbing.
2. Phase II: MOS Build – Populate Data Fields for the Optimization Model.
3. Phase III: Construct and Complete the Optimization Model.

Phase I is the focus of this thesis. The Linear Program (LP) or Non-Linear Program (NLP) that will eventually complete this process will consist of data, variables, an

objective function, and constraint set sections. We will define a preliminary optimization model in this thesis and capture the necessary data elements. I will also summarize a great deal of the constraint set. The eventual optimization model should consist of and resemble the following:

INDICES and SETS:

i	ZIP Code of interest (00010…99985) [1,…,10^5]
j	MOS of interest (00B…98Z) [1,…,264]
k	Reserve Center (The current number of RCs) [1,…,829]

PARAMETERS:

$max\_recruit\_zip_i$	Maximum number of recruits obtained at Zip i [1]
$max\_recruit\_zip\_mos_{i,j}$	Maximum number of recruits obtained at Zip i of MOS j [2]
$target\_mos\_rc_{j,k}$	Target MOS j at RC k
$zip\_rc\_dist_{i,k}$	$\begin{cases} 1 & If\ \ Zip\ \ is\ \ within\ \ 75\ \ miles\ \ of\ \ RC \\ 0 & o/w \end{cases}$
$zip\_rc\_time_{i,k}$	$\begin{cases} 1 & If\ \ Zip\ \ is\ \ within\ \ 1.5\ \ hours\ \ of\ \ RC \\ 0 & o/w \end{cases}$
$weight\_unit_k$	Weighting (priority) of unit at RC k assigned by OCAR [tier 1 = 1, tier 2A = 2, tier 2B = 3, tier 3 = 4, tier 4 = 5, tier 5 = 6]
$weight\_mos_j$	Weighting (priority) of MOS j assigned by OCAR [Top 15 = 1, 2, …, 15; All others = 16] [3]
max_flow	Maximum Flow from any ZIP-RC arc

VARIABLES (*Note: All variables are non-negative):*

$FLOW_{i,j,k}$	Flow from ZIP Code i to MOS j to RC k
$ZIP\_RC_{i,k}$	$\begin{cases} 1 & If\ \ Zip\ \ is\ \ in\ \ RC\ \ market \\ 0 & o/w \end{cases}$
$FILL\_MOS\_RC_{j,k}$	Fill of MOS j at RC k
$OVER\_MOS\_RC_{j,k}$	Number personnel over 100% fill of MOS j at RC k
$UNDER\_MOS\_RC_{j,k}$	Number personnel under 100% fill of MOS j at RC k

FORMULATION:

$$\text{MIN} \quad \sum_k WEIGHT\_RC_k \left(\sum_j WEIGHT\_MOS_j * UNDER\_MOS\_RC_{j,k} \right)$$

$$\text{s.t. (1)} \quad \sum_{j,k}\sum FLOW_{i,j,k} \le MAX\_RECRUIT\_ZIP_i \qquad \forall\ i$$

$$\text{(2)} \quad \sum_k FLOW_{i,j,k} \le MAX\_RECRUIT\_ZIP\_MOS_{i.j} \qquad \forall\ i,j$$

$$\text{(3)} \quad \sum_k ZIP\_RC_{i,k} \le 1 \qquad \forall\ i$$

$$\text{(4)} \quad FLOW_{i,j,k} \le ZIP\_RC_{i,k} * MAX\_FLOW \qquad \forall\ i,j,k$$

$$(5) \quad ZIP\_RC_{i,k} \leq ZIP\_RC\_DIST_{i.k} \qquad \forall\, i,k$$
$$(6) \quad ZIP\_RC_{i,k} \leq ZIP\_RC\_TIME_{i.k} \qquad \forall\, i,k$$
$$(7) \quad \sum_i FLOW_{i,j,k} = FILL\_MOS\_RC_{j,k} \qquad \forall\, j,k$$
$$(8) \quad FILL\_MOS\_RC_{j,k} - OVER\_MOS\_RC_{j,k} + UNDER\_MOS\_RC_{j,k} = TARGET\_MOS\_RC_{j,k} \qquad \forall\, j,k$$

[1] *max_recruit_zip*$_i$ ➔ f(demographic factors)
[2] *max_recruit_mos_zip*$_i$ ➔ g(demographic ZIP Code factors)
[3] May consider regionalization of MOS priority

Constraints 1 and 2 above are formulated by using the methods of this thesis. Variable construction in this manner provides control of the MA population in the ZIP Code. Note that some ZIP Codes are larger than others. The objective function minimizes the shortages of personnel by MOS, weighting each MOS, and weighting RCs by priority. The optimization distribution model depends on the outcome of the findings of the MOS Build in Phase II. The outcome of the specific MOS analysis will determine the actual model form. Programming the constraints achieves the following:

1. Limits the number of recruits per ZIP Code to its maximum level;
2. Limits the number of recruits in a given MOS per ZIP Code to its maximum level;
3. Limits each ZIP Code to at most one RC or a separate ZIP Code distribution plan to share market ZIP Codes (this feature can be relaxed);
4. Forces flow from a ZIP Code outside its allowed RCs to zero;
5. Excludes ZIP Codes from RCs that are too far (distance);
6. Excludes ZIP Codes from RCs that are too far (time);
7. Balance equation showing personnel assigned by MOS in an RC;
8. Balance equation for Fill, Target, Over, and Under constraints.

This thesis determines the bounds for the constraints of type 1 and 2.

This formulation assigns ZIP Codes to RCs. A subsequent formulation will assign RCs to a given ZIP Code, and the other ZIP Codes to that RC. By changing the units assigned to a given RC, $target\_mos\_rc_{j,k}$ changes. This allows exploration of different assignment of units to existing RCs. This, too, can be explored in Phase III.

Unfilled unit positions hurt the readiness and training of USAR units. Unit positioning with respect to the population has also been a long-term problem. Finding an adequate number of high-quality recruits has also been a problem for units with positions requiring higher MOS ASVAB line scores. The development of a unit positioning and quality assessment tool will greatly assist unit fill and retention rates. This three-phase model will provide insights and help solve one of the most complex problems facing the USAR. It will involve the development of several tools and analyses. Once complete, it will greatly improve OCAR's ability to manage the reserve force.

2. Source

Finding a single cause of TPU unfilled vacancies is very difficult. Historical fill and retention rates of USAR TPUs in their respective geographical locations may give insight as to potential reasons. To study the system we need to determine factors associated with inability to fill TPU vacancies. There are several reasons for the inability to fill the units, and unit location may prove to be most significant.

Figure 1.1 shows the actual USAR TPU locations. There are 829 Reserve Center (RC) stations housing more than 4,200 units. Historically, a unit's actual geographical location is associated with unit fill rates (USAREC, National Market Analysis (NMA), 2000). Not having sufficient numbers of qualified military recruits available in a market (population) definitely influences the fill rate of a unit and its readiness.

The USAR currently adopts a policy of relocating units having fill problems by use of Market Supportability Studies (MSSs) provided by USAREC. This has proven beneficial over time. As the USAR relocates units into better markets, unit fill rates have increased. However, the MSSs provided by USAREC consider only the volume metric for the population. This analysis considers not only the volume but also market quality and vocation.

USAR TPU Locations

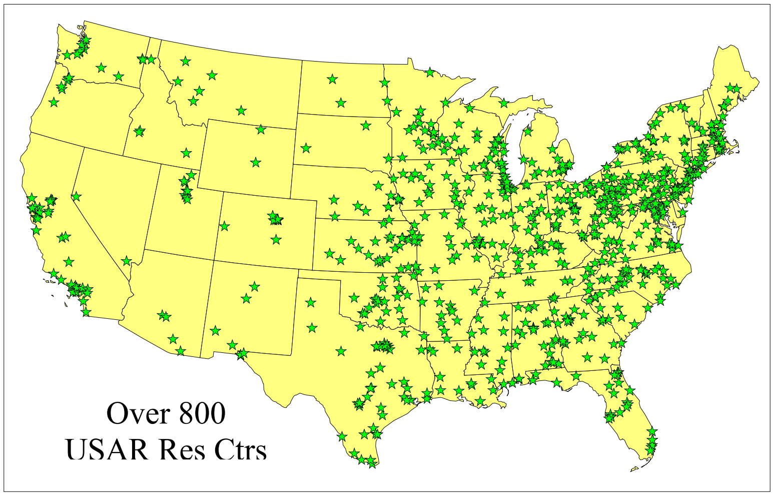


Figure 1.1: USAR TPU Locations

The overlay of Figure 1.1 and Figure 1.2 demonstrates that USAREC recruiting station locations are often in close proximity to TPUs. Each unit has many MOSs USAREC attempts to fill. The national fill priority for MOSs takes precedence over

USAREC Recruiting Station Locations

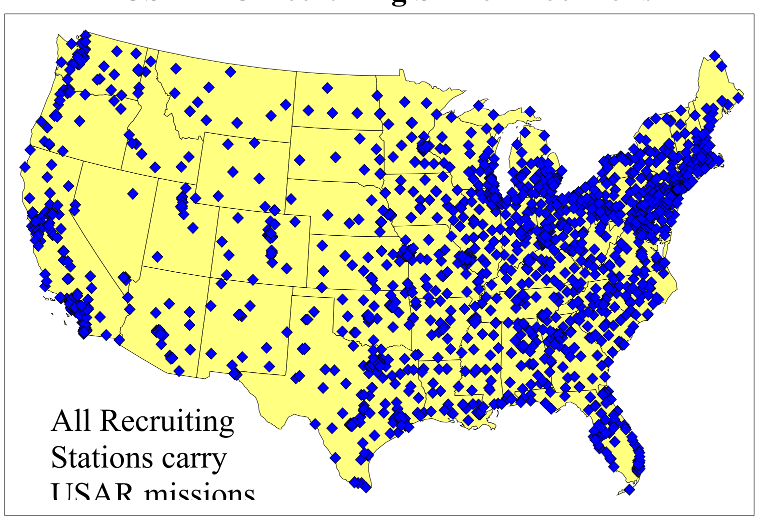


Figure 1.2: USAREC Recruiting Station Locations

locally needed MOSs. Some of the problems causing poor fill rates may be TPU attrition, recruiting difficulties pertaining to unit stationing and resources, the draw-down

of US Forces, local economic situations, structure changes associated with changing missions, TPU deployments, and the competition associated with population vocational availability. Other problems include education and skill training availability, job market, economy, unemployment rates, and sister service competition.

Figure 1.3 demonstrates that TPUs are properly located near the MA population. These are Hot Spot Projection maps. They are thematic mappings of population information buffered on aspect intervals (75 mile radii) using the 2000 US Census data for the MA population.

These maps demonstrate coverage and stationing of both USAR TPUs and USAREC recruiting stations with respect to the markets. The upper left graphic shows the actual placement of USAREC recruiting stations, while the upper right graphic shows the actual USAR TPU placement in the market. The lower graphic is an overlay of both the stations and TPUs with respect to the markets. This graphically demonstrates the recruiting coverage for the TPUs

With a few exceptions, Figure 1.3 strongly suggests the recruiting stations are properly aligned with TPU locations in the market. USAREC Marketing personnel carefully review these exceptions and make minor adjustments to station recruiting missions for TPU coverage. This information and the manner in which USAREC conducts its mission and market planning to provide coverage for the TPUs, along with provisions for high priority TPUs, suggests that TPUs are located with respect to the market.

Although unit locations appear to be aligned with the population, it is possible that TPU force structure may be misaligned within their respective markets. Looking at the vocational aspects of the market may shed light on this consideration. The type of employment available in geographical locations affects personnel availability for unit fill. The analogy for the argument is that if a steel manufacturing plant is to be built in a particular location, it requires sufficient personnel, within commuting distance and with certain vocational skills, to operate the facility. The unit fill potential is the extent to

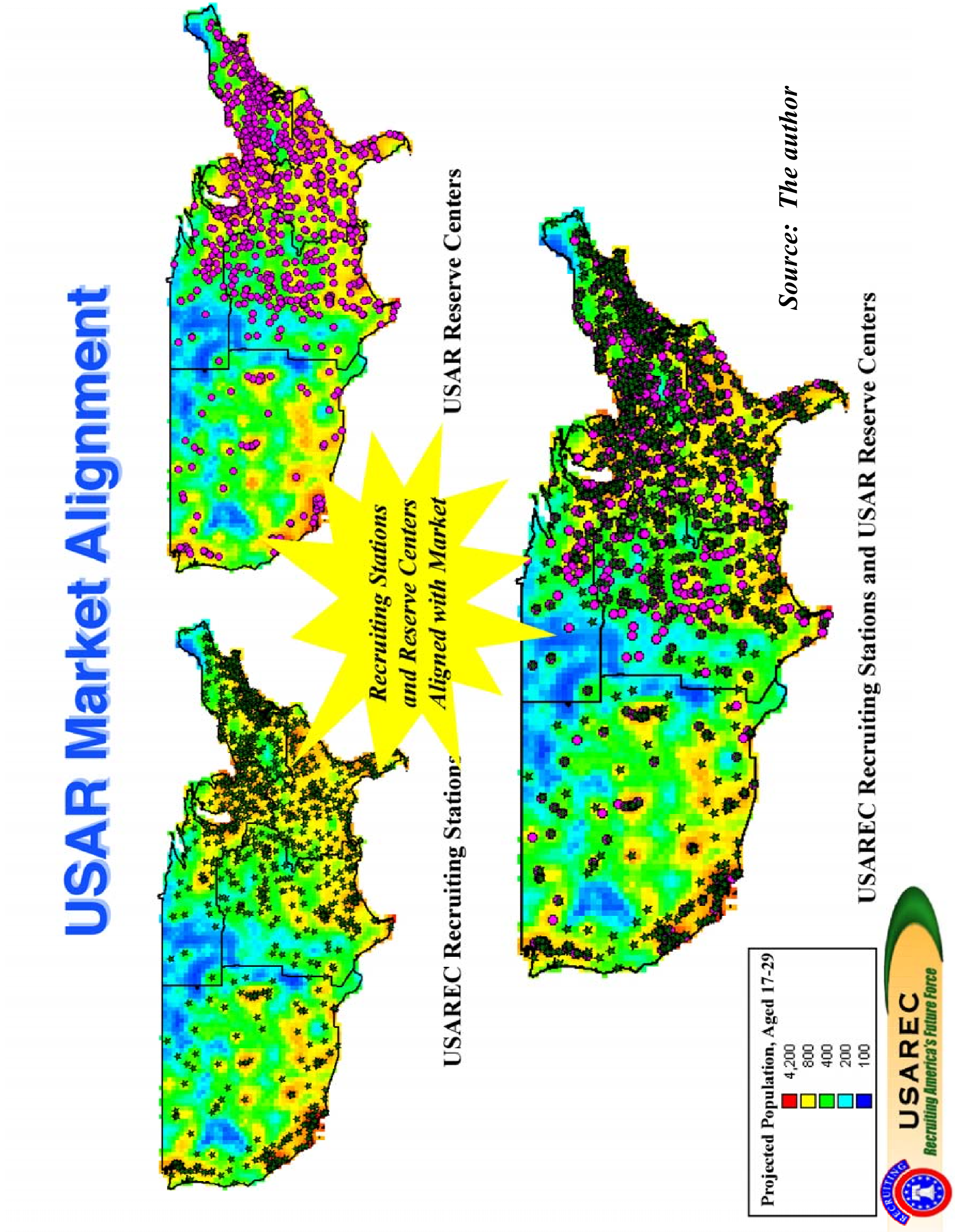


Figure 1.3: USAR Market Alignment – Hot Spot Projection Map

which a unit can expect to find the requisite number of skilled personnel in the local market. Because a unit should be close to the supporting population, the Army targets the recruitment of personnel to fill a unit based on the MA population within 75 miles or a 90 minute commute. Recruits may join a unit outside this range, but this is an exception to policy rather than the rule.

The vocational support available to fill a unit's vocational requirements can be determined by matching the unit's MOSs to the local workforce's vocational availability. This latter information is available from the Bureau of Labor and Statistics (BLS) by ZIP Code. With the BLS data, we can identify market vocations. We can specify the top eleven vocational aptitudes of the ZIP Code. We can then ascertain if there exist sufficient quantities of personnel available to fill unit vacancies.

This is the reasoning behind the unit force structure breakout and stationing. A battalion may not be successful at a particular location, but a smaller company or platoon might. Regulations require the USARC to submit any proposed stationing actions or changes to USAREC. USAREC is then responsible for conducting a Market Supportability Study to ascertain the current force structure and determine if there is sufficient MA population to support any changes.

Another tool assisting in this process is the Competitive Market Analysis – Reserve (CMA-R). The CMA-R reports the local market availability of US Army and sister service competition at an RC or other market levels. This tool enhances the USAREC's ability to assist in market analysis by demonstrating what potential, if any, exists in the market.

It may be beneficial to place our organizations in locations where the organization's vocations are similar to those in the market. For example, assume we have a total of 1000 MA personnel for a particular RC, of whom 130 are identified as transportation workers. Suppose further that two local trucking firms employ 150 over-the-road and long-haul transportation workers. Rhetorically, where do we locate our units to draw on the market vocations?

Would a Transportation Battalion (Medium/Heavy Transport), requiring 630 personnel of whom 475 are actual truck drivers (MOS 88M) be successful in this

particular area? We need to know if there is other useful information available to the TPUs and Regional Support Commands (RSCs). Based solely on volume results of the MSS, we might conclude that it cannot be supported. But with the knowledge of local market vocations, our conclusion could be different. Knowing market vocations may assist in positioning units in those markets.

Modeling the process of filling unit vacancies will greatly assist the recruiting efforts and TPU fill rates. There are several tools available to assist in unit fill. Existing tools are the NMA, MSS, and the CMA-R. The USAR cannot begin to be ready and trained without sufficient personnel. Determining factors associated with unit fill, unit positioning, quality assessment, and demographic considerations for potential success in meeting force structure objectives is the first step in achieving ready, trained, and deployable units.

We want to position RCs to support recruitment for them. We hypothesize that recruitment is affected by demographics, vocational aptitude, and economy of the surrounding area. We want to model the recruiting potential by MOS and ZIP Code so we can enter this information as a constant in the optimization distribution LP model. To model recruitment potential by MOS and ZIP Code, we must mine several large incompatible databases to construct our data set.

This data mining is an enormous task. We accomplish it, automate it, and document it. Using our data set, we illustrate the recruit potential model for 4 key MOSs. A second thesis can complete the recruit potential model for the other 260 MOSs and analyze the model set for commonalities and distributions. A third thesis can implement the full LP model and develop the optimal RC unit distribution plan.

THIS PAGE INTENTIONALLY LEFT BLANK

II. SUPPORT, ISSUES, AND COURSE OF STUDY

A. SUPPORT AND POSSIBLE CAUSES

Analysis to support USAR TPU fill is ongoing. There are other tools used in conducting market research, operations analysis, and subsequent analysis of these items of interest. USAREC provides additional support for market analysis in other forms throughout the year. Some of the support includes the following: NMA, MSSs, CMA-R, Demographic Support (USAR Enhanced Applicant File), Market Research Tools, Market Estimates, Population Studies, Unit Attrition Studies, etc. If USAREC and the USARC do a good job in supporting the RSCs and TPUs, what is the cause of the unit fill problem experienced by some TPUs?

The fundamental problem appears to be determining causes for the unit fill problem. Within this scope, how do we determine the appropriate markets for TPU structure? Trying to define "appropriate" among 10 RSCs and over 4,200 TPUs is challenging. What is considered appropriate for one may not be appropriate for the other.

Previously, we saw Figure 1.1 depicting the actual unit locations of the CONUS USAR TPUs. There are significantly fewer than 4,200 TPU locations because multiple units can be housed at one location. Cost of facilities is a key factor. Therefore, many RC has multiple units stationed at its location. They may be grouped because they are similarly typed, have the same higher headquarters, have a similar mission area, etc. Army Regulations require unit stations be shared among several organizations. There are other factors influencing the outcome of unit stationing actions.

Other influences include, for example, historical and political boundaries. Examples are units traditionally located in areas such as Philadelphia, Boston, or some other area of historical significance. Some politicians firmly believe their constituents want to have units stationed in their legislative districts because "the unit has always been here" or the local economy needs the payroll.

B. NON-DEMOGRAPHIC ISSUES AFFECTING UNIT FILL

1. Considerations

In this section, we list some issues affecting unit fill not included in the analysis of this thesis. Although we do not have data to conduct an analysis on the effects of enlistment inducements, it is important to mention them as part of the discussion of unit fill. Incentives may affect an applicant's decision to join a unit when his original inclination was not to join or he wanted to choose another MOS that may not be available in a particular RC.

A small discussion follows on policy options for the CAR to provide enlistment incentives to better penetrate and acquire the skills of the market. We will refer to this as regionalization. Regionalization also affects the market. Providing bonus or monetary incentives to the population is an enticement to enlistment. We use enlistment bonuses to entice recruitment.

MOS bonus and educational incentive programs greatly affect unit fill. Offering incentives supports the national fill requirements by MOS. But unit geo-demographic considerations may have not been supported. It may prove beneficial to localize incentive programs thereby supporting the local commanders' ability to offer bonus and incentives to fill particular MOS requirements not listed as part of the national priority of needs. For example, say MOS 88M (Transportation Specialist) is listed as one of the national priority MOSs, the top fifteen undermanned MOSs, to fill because of the collective fill rate of the MOS. However, it may not be the MOS needing to be filled in a particular region of the country. There may be a requirement to fill MOS 63B (Light Wheeled Vehicle Mechanic) in this area. It may prove beneficial to offer an incentive or bonus program for 63B as opposed to 88M is this particular region.

Educational incentives may not be quite enough to convince an individual to join a unit for a particular needed MOS. However, having a regionally needed MOS associated bonus may be enough enticement for the same individual to enlist for the particular needed specialty. Otherwise the USAR might lose the individual to a sister service component which can satisfy the individual's interest in a particular specialty.

Regional impacts are significant when considering the long-term effects of unit fill on readiness. There are several national programs and activities affected when USAR units are not filled. They include:

PROGRAMS	ACTIVITIES
POM Projections	Deployment Capabilities
OMAR Funding Resources	Unit Readiness
CAR's funding and resource allocation level for successive fiscal years	Media Attention
	Force/Power Projection Capabilities
Enlisted Incentive Programs	Unit Leadership
Educational Incentive Programs	Training

The dilemma is what to do about the regional performance of USAR TPUs. TPUs have the responsibility to train for war. Their preparedness is instrumental to the success of this nation to achieve its goals. Prioritization is paramount to achieving fill rate success. Priority units have fill priority. The following two areas need consideration as well:

1. Regional needs by Area Support Group (ASG) or some other methodology.
2. Incentive and bonus needs by ASG or some other methodology.

2. Demographics and Unit Positioning Effects on Fill Rates

The rationale for conducting this study is based on the principle of local demographic effects. Size, type, employment, vocations, education, and other factors affect local markets. Recall that the USAR has a geographical constraint limiting its market draw to the population within 75 miles or a 90 minute commute.

We will demonstrate the affects of demographics. We hypothesize that the local employment or unemployment rate has an effect on the fill rates of units.

Force structure composition in local markets is important to unit fill. We can see these effects if the population majority, in a particular area, is more likely to join a maneuver unit than a transportation unit. If the USAR places or has transportation force

structure in this area and the ARNG has armor or infantry force structure in the same location, transportation unit fill could suffer.

Demographics and market composition must be addressed when deciding what force structure to place in a particular market.

3. Deployment Tempo Inclusion

The USAR TPU deployments have been on the rise in the last decade. Statistics indicate deployments are up 25% in the past decade. The USAR is being used at an increasing rate. However, at the time of this analysis, it was not feasible to obtain deployment data of USAR units. Deployment effects may not be seen until a few years after the unit redeploys to its home station. Further study in this area may reveal some peculiarities not yet discovered. Consideration of this topic should be included in further studies related to aspects of the unit fill problems.

C. OBJECTIVES

The overall project objective is to establish an optimization model for unit distribution by which to maximize unit fill in markets. The scope is limited by the ability to predict, forecast, or otherwise optimize the unit placement with respect to the population composition. The scope of this thesis is to define the process, define the optimization model, collect the data elements, and scrub these elements. This information will feed subsequent phases of the project, especially Phase II. Recall that Phase II establishes the constraint set of the optimization distribution model to complete the analysis.

The goal of this thesis is to identify the supportability of TPUs by the size and quality assessment of the population. To do that we draw the appropriate data, summarize the data, and analyze current unit structure with respect to population supporting USAR unit fill rates in their current markets. We will establish whether current locations can support certain MOSs. We will accomplish this through regression analysis by modeling of the number of expected contracts from each ZIP Code and the

expected number of contracts by MOSs each ZIP Code can support. The response variable will be contracts. The predictor variables will be the BLS vocational inclination data groups (11), MA population (1), Microvision 50 (MV50) Lifestyle segmentation categorized by groups (11), quality assessment via ASVAB scoring (10), quality assessment via Armed Forces Qualification Test (AFQT) (1), and unemployment rate (1) for each ZIP Code.

We will focus on the efforts of the USARC and USAREC to accomplish their annual USAR enlisted accession mission. Specifically, we address the current TPU vacancy problem and the unit positioning or stationing problems. We will examine and understand some of the basic concepts associated with identifying the problem, arriving at a feasible solution, and communicating this information to the appropriate decision maker for action.

USARC's Force Structure analytical personnel are the audience for this thesis. Structure positioning with respect to market is one of the keys to success in filling unit vacancies. The right type of unit needs to be in the right market.

We will determine and recommend to the Chief, Army Reserve (CAR) a more appropriate distribution of ZIP Codes to RCs so the current and projected markets can support the TPUs at their respective locations.

D. COURSE OF STUDY

We use regression techniques to maximize the fill rate of USAR units. This regression uses predictor variables including BLS vocational aptitudes of US population, MA population, ASVAB Lines Scores, AFQT Scores, and MV50 segmentation information to gain insight to better unit stationing. We also seek to uncover better practices in stationing actions for USAR units. We would like to answer the following questions:

1. Is there a methodology that enables the USAR to better station units with respect to the population demographics?

2. Is there a significant correlation between unit fill and the vocational propensity of the market or ZIP code of interest?

3. Is there a significant correlation of local market competition factors such as job market unemployment rates, sister service human resource competition, and USAR ability to fill units in these areas?

4. Does the market have sufficient population to meet structure or quality requirements necessary for a particular unit?

5. What insights arise from analysis of the top or most prominent vocations in each market?

6. What are the policy implications for the Chief, Army Reserve (CAR)?

7. What is the effect of relaxing or tightening the commuting constraint?

We explore and evaluate unit positioning with respect to geo-demographic considerations of respective recruiting markets. We identify and subsequently ignore those political encumbrances with respect to historical placement of some reserve units and the constituent population. Historical accessioning information and other relevant data determines the unit fill rate.

We restrict modeling efforts to those methods involving linear transformations, regression applications, forecasting, and optimization techniques that give insight to significant relationships of unit positioning in a geo-demographic market. We will describe the equation of the "top" five MOSs with respect to the variables of interest. The collection of information must be at Zip Code level of detail to create a model to distribute this information to an RC. There is a multitude of information needed to determine the suitability of the MOS in the market. Major data elements used in the analysis include:

a. US Postal Service ZIP Code Master File

b. Bureau of Labor and Statistics (BLS) Vocational Master File

c. Fill Rates of USAR units by ZIP Code or market

d. Force Structure File

e. Local Area Unemployment Master Data File

f. FIP Code Master Data File

g. MOS Quality (QUALS) Master Data File

h. Sister Service (Reserve) Accessioning Data

i. All Army Accessioning Data

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA AND METHODOLOGY

A. DATA SOURCES

Data is essential for analysis. Although obtaining a data set sounds simple, putting the data into a useful format and ensuring it is free from obvious errors was the most complicated part of this analytical process. Placing this data into a useful form is an art as well as a science. All acquired data in this thesis was obtained without monetary expenditure on the part of the analyst. This in itself is a major feat. Appendix A (Table Definitions Dictionary) contains the obtained data on unit stationing, population statistics, force structure files, MA population, and vocational aptitudes of the entire US market by ZIP code or Federal Information Partnership (FIP) code. The FIP code is the state and county origin of the data sampling. Appendix A describes:

a. US Postal Service ZIP Code Master File (http/zip4.usps.com/zip4/zip_responseA.jsp);

b. USAR Force Structure File (FRC_FILE);

c. USAREC Military Available Population Data (PM03);

d. Microvision 50 Lifestyle Segmentation Data (MV50);

e. All Army Accession Data (ALLARMY);

f. Sister Service Accession Data (SISSERV);

g. Qualifications Data (QUALS);

h. BLS Vocational Master File (P050);

i. BLS/USBC Local Area Unemployment Data – County (LAUCNTY);

j. BLS/USBC General Population Employment Data (gp.data.1.AllData);

k. BLS/USBC General Population State Code Data (gp.state).

B. DATA COLLECTION

Appendix B (Table Data Fields and Descriptions) contains information on the tables used in the analysis. There were a number of sources used to obtain data. The actual data collection and preparation consumed more than two months of effort. The author was involved in the initial data-warehousing project at USAREC. This enabled faster data acquisition of the MA population, contract and accession, sister service, and segmentation information used in the analysis. Data warehousing greatly assists in reducing the amount of time required to obtain data elements for analysis. Data elements required about two weeks to acquire once the query for the data was formulated. Query formulation took approximately three days to accomplish. Without the data-warehousing capability, this data collection would have taken over two months to accomplish.

While waiting on these elements, we had to find the vocational information and obtain access to this information by ZIP Code. The author's spouse is a Field Representative for the United States Bureau of the Census (USBC), and helped. This data was obtained by tracking the information back through the Current Population Survey (CPS). These elements took was approximately 3.5 weeks to collect. Once obtained, it had to be manipulated from its source into a workable format for integration into final tabular form taking another three days or so. Total time invested was approximately one month.

Two other hard-to-acquire data sets are the Local Area Unemployment (LAU) county (employment and unemployment) data and the United States Postal Service (USPS) ZIP Code information. The unemployment data is collected and summarized by FIP Code, not by ZIP Code. Once located, this table was copied from the BLS website in text clipping format, as no file transfer protocol (FTP) site was available. Once clipped in text form, we had to find and acquire a way to break the data into useful pieces of information, using a dictionary. We obtained one from the BLS. Once obtained, we used the data dictionary to segment the data into its useful pieces. There are over 2600 counties in CONUS. A great amount of effort was put into to locating a ZIP Code to FIP Code table.

The author recalled a five-year-old table having exactly what was needed. However, since the Postal Service changes ZIP Codes frequently, the data had to be checked and scrubbed for accuracy. The Postal Service currently has over 33,000 listed ZIP Codes. This includes all US possessions and territories. Since our concern was CONUS, this narrowed our scrub to approximately 30,000 ZIP Codes to verify. Since the USPS changes ZIP Codes frequently, the only manageable way to accomplish the ZIP Code verification was to conduct these verifications on-line through the USPS website.

The initial scrub confirmed over 27,000 ZIP Codes leaving about 3,000 to check and verify by hand. This was a tedious task to accomplish. This process took approximately 3 minutes per ZIP Code, working on-line through the USPSs website. The complete task took 150 hours. If we had been able to purchase current the ZIP Code Master File, we might have been able to cut this task duration time in half.

Once the second scrub was complete, we had to resolve by hand over 700 ZIP Codes that were not available on the USPS website. However, I considered them critical for the analysis because the number of contracts produced by these ZIP Codes was greater than 5 per year. If not considered, we could have lost approximately 4,000 annual NPS contracts, out of an average annual USAR accession mission of 20,000 NPS. This process took about 3.5 weeks.

Once we accomplished all these collection tasks, approximately two months had lapsed. As the data arrived, it was necessary to review and become familiar with it. Some data arrived without data dictionaries or other helpful items to understand the tabular contents. Once received, I noticed that some informational items requested did not arrive in a proper format or were not included in the data sent by the provider. Calls and e-mails were made to verify data elements and items not included, taking over two weeks to accomplish.

Some peculiarities found in the data were: no labor force information for some ZIP Codes, no annual production for some ZIP Codes (result of changing ZIP Code data), incorrectly coded information, non-existent ZIP Codes, incorrectly classified lifestyle segmented data, etc. These were addressed to in the development of the final data table containing the ZIP Coded assemble information.

Information arrived in varying formats. Formats included varying text file formats, spreadsheet files, data files, mainframe files, and varying database formats. All the collected information had to be finalized into one table containing all pertinent items with respect to each ZIP Code. The following sources were used in this analysis.

1. United States Army Recruiting Command (USAREC)

USAREC provided data on USAR accessions, listing each applicant's Military Examination and Entrance Processing Station (MEPS) testing data, demographic data, and market segmentation information. USAREC also has in its repertoire of data the useful MA population (*PM03*) derived from commercial source, Woods and Poole. This data was obtained with the assistance of MAJ Michael Kamei and Mr Rodderick Lunger, Programs Analysis & Evaluation Directorate, Headquarters, USAREC, Fort Knox, KY.

The market segments were obtained from a commercial source as well. The clustered data, ZIP+4, were derived from MV50 segmentation data. This data contains 50 market segments characterizing demographics, purchasing habits, etc. This data, along with the Army's accessions data, spans from FY99 through end of FY03. USAREC also provided Sister Service data for the same time period. This data was obtained with the assistance of Mr Rodderick Lunger at (800) 223-3735 (x60358), Programs Analysis & Evaluation Directorate, Headquarters, USAREC, Fort Knox, KY.

2. United States Bureau of the Census (USBC)

USBC provided data on the vocational aptitudes of the entire working population listing each ZIP code's actual vocational inclination using the *P050* Tables from the USBC. We used the Current Population Survey (CPS) data to check the counts of the population and unemployment, and to cross verify the Military Available (MA) population from USAREC data. This data includes the 2000 Census and updates from the Current Population Survey (CPS) data for FY2002. This data was obtained with the assistance of Mrs Susan Fair, Field Representative, USBC; Mrs June Grillo, Senior Field Representative, USBC; and Mr Jamey Christy at (818) 904-6393, Regional Director, US Bureau of the Census, Los Angeles, CA.

3. United States Postal Sevice (USPS)

The analysis used the Master ZIP Code information from the USPS's website, http://www.usps.com/zip4/citytown.htm. The MA population from USAREC PM03 table was cross-verified using the USPS website.

4. United States Bureau of Labor and Statistics (BLS)

The BLS website, www.bls.gov, provided data on the employment statistics of the entire working population, by each FIP Code, listing the actual employment information using the General Population (GP) Tables by county and state from the BLS. The Current Population Survey (CPS) data was used to check the counts of the population, unemployment, etc. It was also cross-verified using the Master ZIP Code information from the USPS's website. The data obtained from the USPS and BLS websites was in text clipping format. It was imported and manipulated using Microsoft FoxPro software into tabular database form for use in this analysis.

5. Office of the Chief of the Army Reserve (OCAR)

Additional data pertaining to USAR force structure (Force_File), recruiting and accessioning priorities, fill priority, and USAR data descriptions were provided by Major Ward Litzenberg at (703) 601-3527, Programs Analysis & Evaluation Directorate at Office of the Chief of the Army Reserve (OCAR), Arlington, VA.

C. DATA PREPARATION

The data preparation took approximately 2.5 weeks to accomplish. Much manipulation, formulation, etc. had to be accomplished to get all the data elements into a common, useful, and usable format for integration. Several software packages accomplished the data preparation aspect of the analysis. The software used to organize, classify, assemble, derive, aggregate, and analyze the data was: Microsoft FoxPro 2.5 (MAC OS), Microsoft Visual FoxPro 6.0, Apple's Text Edit (MAC OS), Microsoft Word Pad, Microsoft Excel (MAC OS), Minitab 10.0 (MAC OS), S-Plus 6.1, and SPSS Clementine 8.0, a data mining software application. Microsoft FoxPro and Clementine

8.0 produced the classification and integration of the data. FoxPro manipulated most of the data tables into a usable format. Once we created the usable format, we used Clementine to graphically demonstrate the data "flow".

Clementine is a data mining application presenting visual representations of data and their elements. It permits limited statistical and accounting operations. It visually allows the user to demonstrate and select data preparation or certain "mining" of data and its elements to filtering. Data "streams" are groupings of different graphical operations from source to sink.

These operations allow the user to demonstrate certain properties of the data. Operations performed by Clementine are: selecting, sorting, setting, appending, filtering, making distinctions, merging, filling, creating, deriving, and collection operations. Input nodes are circles, output nodes are boxes, operations nodes are hexagons (on fields and records), modeling nodes are pentagons, graph nodes are triangles, and supernodes are stars. A user can choose to place the most frequently used nodes in a "Favorites" palette.

Figure 3.1 demonstrates node classification in Clementine. The nodes are graphically and statistically linked in the editor window. One of the first items to consider was to place the data into a usable format. Clementine and FoxPro enabled the data elements to be selected, sorted, assembled, and scrutinized. The figure shows the node types and varieties. There are source, record ops, field ops, graphs, modeling, and output nodes available for use. These classifications permit the performance of a myriad of operations for data manipulation, computations, modeling, and statistics.

The analysis requires RC and ZIP Code level of detail. ZIP code level data formed the basis for the collection and arrangement of data elements to facilitate the analysis. The *JOBMV50* table contains data from tables assembled by ZIP code. All tables containing ZIP code information were verified using the US Postal Service ZIP Code Master File located at http://www.usps.com/zip4/citytown.htm. The USPS web site verified over 33,000 and re-verified over 750 ZIP codes obtained from the various data sources.

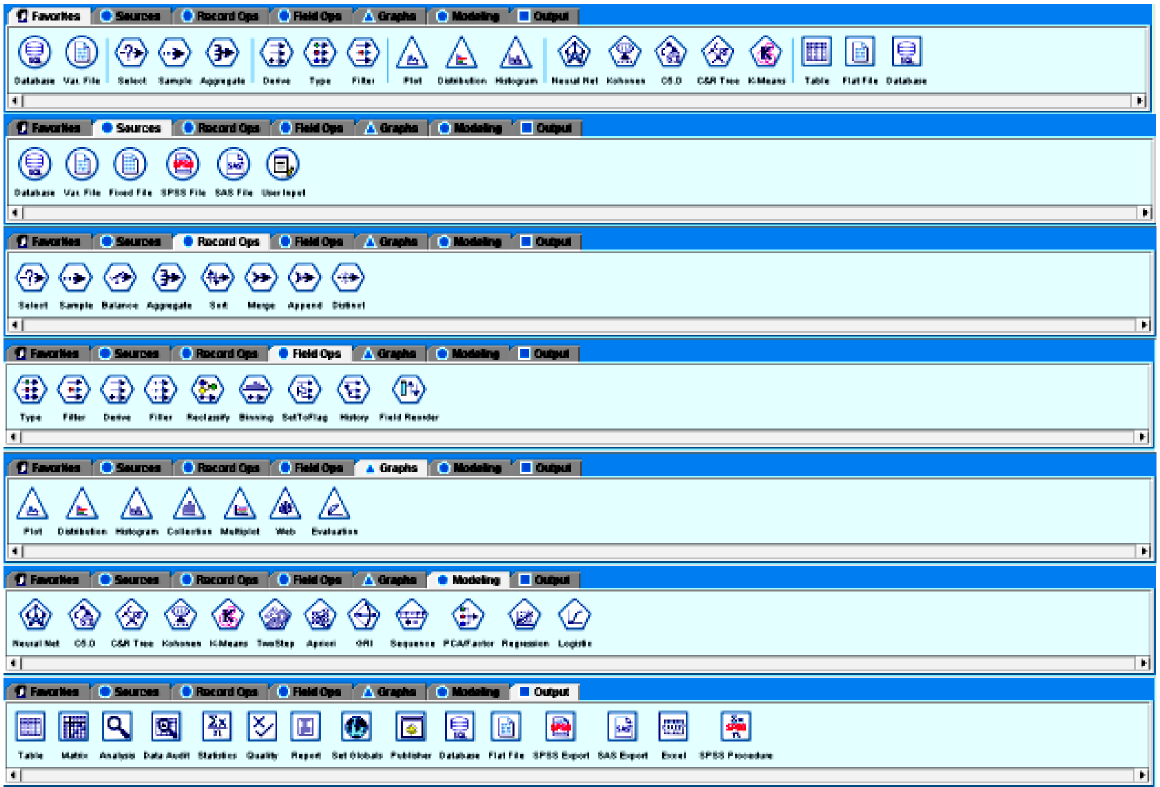


Figure 3.1: Clementine Example Nodes

In Figure 3.2, we combined information, through data manipulation, of the *JOBMV50NEW* with *MAPOPLAU* to create *JOBMVPOP*. *JOBMVPOP* contains BLS vocational, MV50 Lifestyle Segmentation, MA population, and LAU information in one table. Through programming and data manipulation, FoxPro created *JOBMV50NEW* and *MAPOLAU*. *JOBMV50NEW* is combination of BLS vocational and MV50 Lifestyle Segmentation information. *MAPOLAU* is the combination of the MA population and LAU information.

Figure 3.2 demonstrates the results of data mining using Clementine 8.0 software. It shows the kinds of operations used to facilitate data manipulation. The details for the figure are as follows. The INPUT nodes (circular symbols), *JOBMV50NEW* and *MAPOLAU*, are on the left of the graphic. The next nodes (hexagonal symbols), reading left to right, are the TYPE nodes. These nodes confirm the type of data arriving and departing the TYPE nodes. The next two hexagonal nodes are called FILTER and SELECT nodes. They perform the record functions on the data flowing through them.

The other (rectangular nodes) nodes are OUTPUT nodes. These nodes are terminal type nodes. Data flows only into these nodes.

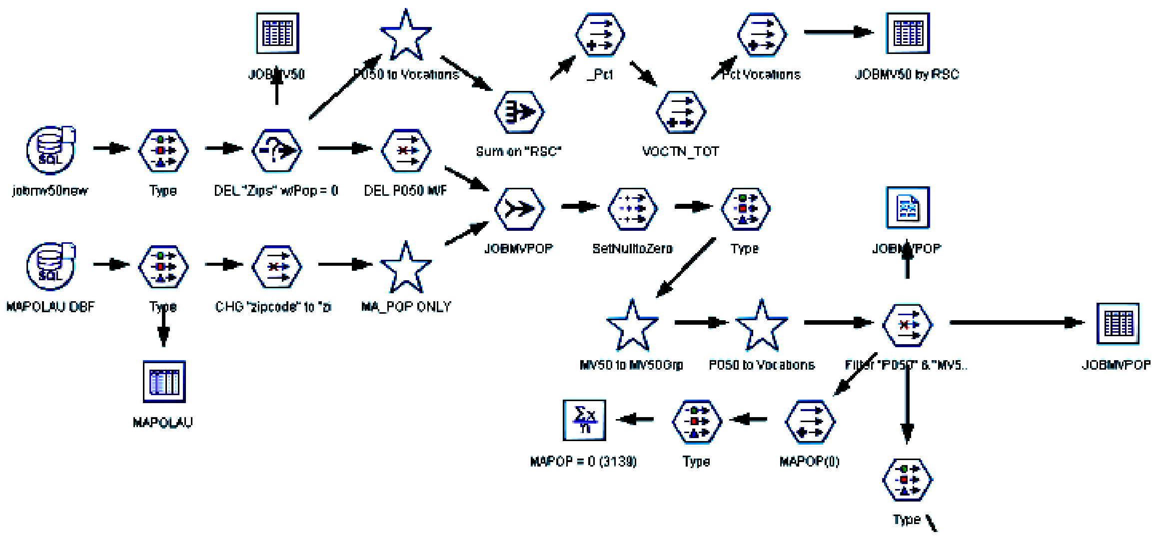


Figure 3.2: Data Mining Using Clementine 8.0 Software

Other nodes depicted in the graphic are SUPERNODES, OUTPUT nodes, and DERIVE nodes. The star nodes are SUPERNODES. They group an informational stream of nodes combining their functions into a single node. Most of the time a supernode use is to denote multiple functions of similar type. It is also used to clean up the graphical flow of data manipulation into one function denoted by the SUPERNODE. The STATISTIC node use is for obtaining certain statistical information about the stream. You can collect information about the stream of data by inserting one of these OUTPUT type nodes. As previously stated, these nodes are terminal nodes. Data only flows into these nodes. The information from the node cannot be used for input into any other stream. The last node depicted in the figure is the DERIVE node. Just as the name of the node suggests, it derives a field or multiple fields from other fields in the stream.

As demonstrated, Clementine is a powerful piece of software which makes data mining very simple and easy to understand. The data flows along the connectors (arrows), called streams. Streams are easily constructed and manipulated. The data flows along the stream paths, from source to terminal nodes, performing operations on

the data resulting in useful information. Looking at the input data and deriving a useful table is beneficial to the analysis in both time and programming effort. Figure 3.2 shows the derivation of the MA population and MV50 segmentation data into the *JOBMVPOP* table. The table is a collection and assembly of data at ZIP code level.

The analysis incorporates the PM03 MA population data. FoxPro 2.5 (MAC OS) and Visual FoxPro 3.0 Relational DataBase Management Systems (RDBMS) were used to bring the information to a useful format. PM03 determines the MA population for each ZIP code. We derive the *MAPOP* from the PM03 table using FoxPro.

Figure 3.2 contains data from the varying sources summarized in the *JOBMV50NEW* (update from *JOBMV50*) table. The two tables providing principal source of information are: *P050* and *MV50* tables. The resulting table, *JOBMV50NEW*, is deemed JOB, from the *P050* table, and MV50, from the *MV50* segmentation data. Also incorporated in the *JOBMV50NEW* table is the *LAUCNTY* table data. This information is the Local Area Unemployment (LAU) data by county for 2002. BLS and the CPS verified this information in 2003. It has the labor force, employed, unemployed, and unemployed rate figures by FIP code.

One additional table supporting the *JOBMV50NEW* table is the *gp.data.1.AllData* table. This table provides the General Population (GP) employment information by FIP code. This table has the average annual historical unemployment rates from 1981 - 1998. It differs from the *LAUCNTY* table, in containing simply the unemployment rate figures for each FIP code along with comments on data specifics.

Obtaining ZIP code detail about our data and population is key to the analysis. Unit authorizations, by MOS, are the basis of the analysis. The USAR *Frc_File* identifies unit authorizations and on-hand totals for all MOSs. Using Clementine, we can choose to include or exclude certain aspects of the data. In establishing the USAR *Frc_File* information, the scope of this analysis excludes the officer and senior enlisted force structure. This is done by the use of select nodes in Clementine.

One item needed for the analysis is the $\text{target\_mos\_rc\_}_{i,j}$. Once we obtain all the demographic information by ZIP code, we can begin other required assembly of the data. The first needed item is Army contract data. We want to determine how many contracts

we obtained from each ZIP code to determine penetration rates of the market. Remember, market is the collection of ZIP codes surrounding the RC within 75 miles. The *RCMKT75* table is the origin of this information. The author created this particular table from RC ZIP Codes. We can determine the units needing personnel fill from the USAR *Frc_File* table. This table has the USAR force structure composition for each unit. For this analysis we will use an extract of the information in the *Frc_File* table called *USARTOT*.

The extract contains the enlisted population, specifically, the skill level 1 and 2 force structure. Our focus is the problematic junior enlisted. Since we have the force structure, we know each MOS required at each RC. If needed, the model can later incorporate all the force structure. Armed with this information, we can use the *QUALS* table to ensure the population scores, on the ASVAB, are sufficiently high enough to qualify for the force structure at its current location.

For example, Figure 3.3 demonstrates the use of Clementine to merge the information contained in the *QUAL* and *USARTOT* tables. During the execution of the MOS Quality Check table, Clementine displays the use of information by turning the input tables purple and the lines linking the data elements green. This shows the graphical representation of the flow of data and the operations performed on the data at each node. Appendix E (Clementine Screen Snapshots) contains details of all constructed streams of data collected, assembled, purged, and extracted.

Here is a summary of the data inputs and derivations. *ALLARMY2* created *ALLARMYCLEAN* and *AllARMY_MOSQualify*. *ALLARMYCLEAN* has all "duplicate", "no ZIP code", and "no AFQT" records stripped from the original data source, *ALLARMY2*. *AllARMY_MOSQualify* is the result of checking the LSCAT against each MOS in the inventory to see if the accession qualified for the MOS. If they qualified for the MOS, we increased the tally for the MOS for the particular ZIP code. The resulting table contains the MOS total qualified for the ZIP code.

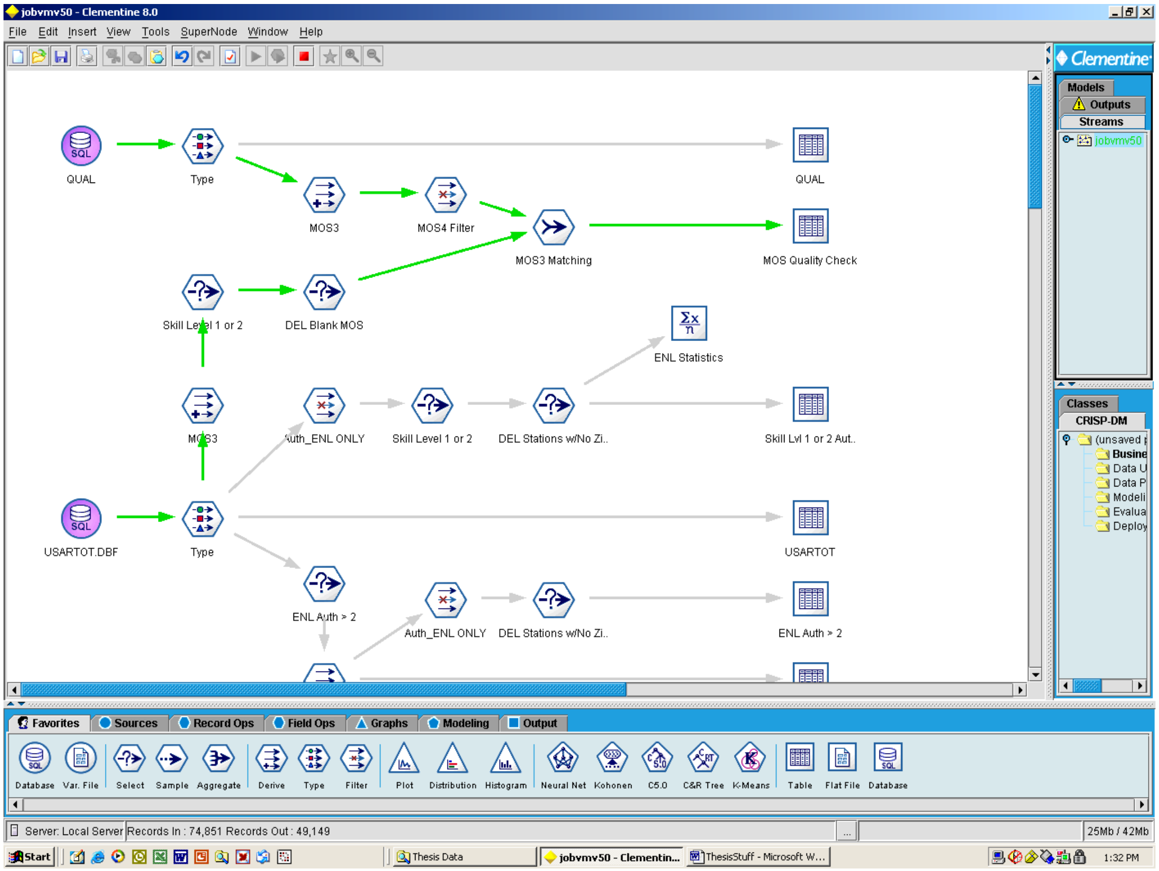


Figure 3.3: Using Clementine to Conduct a Records Merge

We transformed and manipulated *AllARMY_MOSQualify* to derive the necessary information for the analysis. Dr. Samuel Buttrey, Naval Post-Graduate School, using S-Plus code, performed the manipulation of the data to create the tallies for the MOS. We did not carry the column headings for each MOS as they were created since they are in numerical order. After the tallies are complete, we had to place the data back into columnar arrangement to complete the summary of the MOS by ZIP code. C code, programmed by Dr. Samuel H. Buttrey, completed the transformation of *AllARMY_MOSQualify*. The author completed the assembly using S-Plus, MS Word Pad, and Clementine text OUTPUT nodes. We constructed, derived, and assembled the *ARMYbyMOSbyZIP* table using Clementine streams by merging *ALLARMYCLEAN* and *AllARMY_MOSQualify*.

To place the tables into a useful format required the merging of the four individual tables into one. Again, Appendix E contains the details of the merge. We

merged *JOBMVPOP*, *ARMYbyZIP*, *ARMYbyMOSbyZIP*, and *SISERVAFQT*. We previously discussed the details of *JOBMVPOP*. *ARMYbyZIP* contains Army accession data by LSCAT and AFQT for each ZIP code. We previously covered the details of *ARMYbyMOSbyZIP*. Lastly, *SISERVAFQT* has the same information as *ARMYbyZIP*, except *SISERVAFQT* does not have the LSCAT for the Sister Service data. Sister Service data contains data for Marine Corps, Navy, Air Force, and Coast Guard Reserve Components.

Table 3.1 shows a summary of the tabular information associated with the data derivations and manipulations. It contains the file name, number of fields in the file, and the record count for the tables. For example *JOBMVPOP* has 32,873 records and 32 fields: 12 vocational, 12 segmentation, 8 population, and 1 ZIP code fields. *SISERVAFQT* has 30,751 records and 29 fields: 9 AFQT, 19 test score category, and 1 ZIP code fields. *ARMYbyZIP* has 33,178 records and 66 fields: 12 vocational, 15 AFQT, 30 LSCAT, 8 test score category, and 1 ZIP code fields. Lastly, *ARMYbyMOSbyZIP* has 33,124 records and 266 fields: 264 MOS qualifications, 1 count, and 1 ZIP code fields. When merged, these four tables combine into the *ALLDATAbyZIP* yielding the final table for the analysis. This table contains 29,865 records and 392 fields.

FILE	FIELDS	RECORD_CNT
JOBMVPOP	32	32873
SISERVAFQT	29	30751
ARMYbyZIP	66	33178
ARMYbyMOSbyZIP	266	33124
ALLDATAbyZIP	**392**	**29865**

NOTE: The Final ALLDATAbyZIP table is an inner join table containing fewer records than the tables joined (even the minimum number of records – 30,751). I omitted some records with discrepancies and the inner join deleted incomplete ZIP Code information. Thus the Final ALLDATAbyZIP table contains 29,865 complete records.

Table 3.1: Clementine File Creation and Table Derivation Data

This final table, *ALLDATAbyZIP*, represents almost three months of data requesting, collecting, manipulating, assembling, etc. The latter parts, manipulating and assembly would have taken at least three times longer using software languages already

known and understood to the author. The learning curve associated with using and understanding Clementine was about 2-3 weeks.

Clementine greatly assisted in the development of this analysis. The amount of time devoted to getting the data into a usable format is approximately the same as using other software programming languages. However, Clementine is a graphical visual tool allowing a multitude of input formats whereas data formulation and manipulation must be in certain formats to work with database or SQL programming languages. The advantage is these streams of information are already constructed; the data updating can be an automated process without the additional labor and worry of formatting using other software languages.

Appendix E contains the detailed streams constructed in Clementine. The screen snapshots are clearly visible and understood by giving attention to the data streams and the node operations performed on the data. Now that we have seen how to put the data into a useful format, the next chapter develops the analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. THE ANALYSIS

A. POSITIONING UNITS TO MAXIMIZE FILL RATES

The study reveals there are several items of interest with respect to the unit positioning and quality assessment of the markets. Each ZIP code represents a unique contribution to the overall needs of the United States Army.

Our original thoughts were to create data elements for a time series forecast and analysis. We may be able to create a more effective model by using and applying time series forecasting methods. We could accomplish the collection effort; however, our current model has nearly 30,000 ZIP Codes and 432 predictor variables. With six years of information (FY1998-FY2003) times 12 months per year, 72 times more information would need to be collected. Therefore, our resulting data table would be approximately 30,000 by 30,000.

If we were to use monthly time series, our data collection efforts would increase 72-fold making the analysis nearly impossible on a stand-alone PC. The computational effort needed might increase 72-fold or more depending on the processor. Current data streams constructed in Clementine take nearly 42 minutes to run on a 2.80 Mhz Pentium IV processor with 1 Gb of RAM, 60 Gb hard drive, and a LAN access server of over 300 Gb.

We decided to use the 30,000 by 432 table for our contract and MOS regression equations. As explained in chapter 3, understanding the data elements and their relation to the analysis is key. To demonstrate the analysis, we will walk through fitting a model.

I created a table associating MOS to BLS vocations. This information will assist in determining whether the market has a sufficient quantity of this particular vocation to support our force structure. For example, why not locate an engineer construction support company where the prominent vocations of the area are machine operators, craftsman, and laborers?

Using this consideration, we have a rationale to determine force structure placement with respect to the market. Appendix C (Occupations and Working Class

Categories) contains the categorical occupations across the US. This tabular information, from BLS and USBC, contains the most prominent vocations by ZIP code. We develop regression equations for each MOS using this information as predictor variables. We begin to understand why we have a problem. Misalignment of the vocations of the area with the force structure can contribute to poor unit fill.

The next data item used for analysis is the LSCAT information obtained from all Army contracts from FY1999 – FY2003. This information contains the LSCAT scores for each ZIP Code. LSCAT gives information about the quality of the accession. Without it we do not know if we can support the specific jobs in the unit force structure.

Once we have found the MOS regression equations, we can determine which units can be supported by a unit's particular ZIP code. Knowing this information will greatly assist in the constraint set development for the optimization distribution model in Phase III. This will assist in completing the MOS regression equations for Phase II.

B. DATA FAMILIARIZATION

We need to determine the appropriate predictor variables for each model. It is reasonable to assume that population, vocations, lifestyle segmentation, LSCATs, etc. are market influencers. The first question is: How many contracts can I expect to obtain from each ZIP Code? A cursory evaluation of data yields a correlation (0.7737) of the MA population and the number of contracts in the ZIP Code. This is reasonable since contracts should increase as the population increases.

We next examine the data graphically. Figure 4.1 demonstrates the lifestyle segment group percentages for the USAR contracts and the population. There are 11 of these groups plus one segment with incorrectly grouped individuals (MV50GP00&99). This segment grouping was the result of misclassified contracts. The figure shows that some segments are recruited or join proportionally more than other segments.

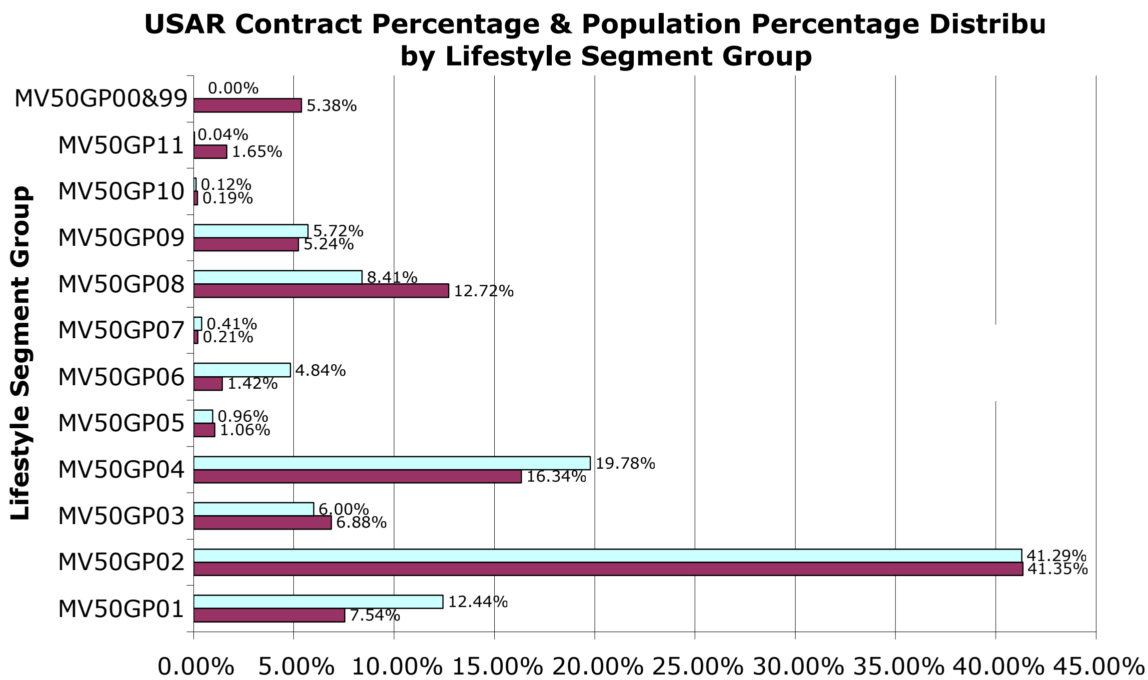


Figure 4.1: USAR Contract & Population Percentage Distribution by Lifestyle Segment Group

Note the distribution of contracts and the population. The distribution of the MV50 Lifestyle Segment Groups for USAR contracts is similar to that of the population, at least in the top 70% of the segment groupings. Segment Group 2 for the USAR is 41.35% compared to 41.29% for the population. Segment Group 4 for the USAR is 16.34% compared to 19.78% for the population. Finally, Segment Group 8 for the USAR is 12.72% compared to 8.41% for the population.

Lifestyle Segment Groups 2, 4, and 8 represent over 70% of the USAR contracts. The distribution of the MV50 Lifestyle Segment Groups for the USAR is similar to the remainder of the Army. It appears as though the USAR contracts a large number of personnel from these three segment groups. Therefore, we expect that these segment groups will be represented in the final regression.

Appendix D (Microvision 50 Lifestyle Segments) contains the segment groupings. Segment Group 2 consists of Segments 10, 11, 16, 17, 18, 22, 35, and 38.

This grouping is composed of families. Segment Group 4 consists of Segments 8, 12, 15, 32, 34, 39, and 40. This grouping is composed of people who are single. Segment Group 8 consists of Segments 24, 42, 43, 44, and 46. This grouping is composed of families as well. A Chi-Square test for the difference of equal proportions shows statistically significant differences. However, when you look at their distributions, they do not differ by much.

It is a reasonable expectation that the vocational composition differs at ZIP Coded level. There may be some kind of grouping that aggregation would show some similarities, at least in the majority or major categories. We explore the data by:

1. Grouping data by FIP Code (over 2600);
2. Grouping data by Metropolitan Statistical Areas (MSA) (over 1300);
3. Grouping data by State (49 – CONUS);
4. Grouping data by ASG (over 20);
5. Grouping data by RSC (10).

We decided to look at a summary categorization by RSC. There are 10 CONUS RSCs and we also had data on the 9th ARCOM. We conducted a Chi-Square test for similarities in the RSCs' and the ARCOM's vocations and lifestyle segmentation. The results indicate the RSCs differ in segments and vocations. We performed the Chi-Square test for similarities on the population raw data. Tables 4.1 and 4.2 are shown in percentages for display only. One would not be able to see the difference with the raw data, so we demonstrate the difference by using the percentages. The actual Chi-Square value for the raw data is located at the bottom of the table.

The tables indicate they are very different in the percent of several vocations and lifestyle segments. The vocational table shows those differences. Some are strikingly different such as FAFOFISH and TRANSPO vocations.

RSC	EXEC MNGE	FAFO FISH	ADMIN SPT	PROF SNL	TECH SPT	SVC OTHR	SVC PROT	SALES	CRFTS MAN	LABOR ERS	TRANS PO
9AR	22.562%	0.787%	7.850%	14.387%	3.003%	15.666%	2.115%	15.910%	8.383%	0.231%	9.107%
63rd	24.593%	0.557%	7.807%	15.421%	2.823%	11.562%	1.596%	14.977%	8.668%	0.267%	11.729%
70th	23.704%	0.965%	7.320%	15.479%	3.132%	11.167%	1.259%	14.442%	9.284%	0.277%	12.971%
90th	22.255%	0.573%	7.507%	14.450%	3.337%	10.730%	1.611%	15.006%	10.806%	0.352%	13.374%
96th	24.214%	0.982%	7.664%	15.036%	2.992%	10.754%	1.243%	15.040%	10.404%	0.356%	11.315%
89th	22.339%	1.185%	7.557%	13.911%	3.509%	10.729%	1.183%	14.639%	9.330%	0.249%	15.371%
81st	21.615%	0.484%	7.379%	13.505%	3.433%	10.646%	1.563%	15.146%	10.581%	0.324%	15.324%
88th	22.396%	0.469%	7.652%	14.419%	3.399%	10.342%	1.311%	14.497%	8.588%	0.212%	16.715%
99th	24.356%	0.315%	7.886%	16.211%	3.471%	10.264%	1.564%	14.131%	8.755%	0.263%	12.782%
77th	25.036%	0.175%	8.153%	16.551%	3.620%	10.965%	2.040%	14.803%	7.281%	0.213%	11.163%
94th	25.959%	0.237%	7.708%	17.392%	3.602%	10.125%	1.422%	14.121%	7.801%	0.240%	11.393%

NOTE: Chi Square Test for similarities conducted on Raw Data, not the Percentages

Pearson's chi-square test without Yates' continuity correction: X-square = 2745863, df = 100, p-value = 0

Table 4.1: Chi Square Testing of Vocational Aspects of RSCs

[The percentage of population vocations for each RSC. This table demonstrates the difference in vocational composition of each RSC.]

RSC	MVGP01	MVGP02	MVGP03	MVGP04	MVGP05	MVGP06	MVGP07	MVGP08	MVGP09	MVGP10	MVGP11
9AR	37.134%	20.527%	6.874%	9.091%	1.319%	1.141%	0.525%	1.013%	22.155%	0.218%	0.004%
63rd	22.663%	28.143%	3.519%	25.740%	0.544%	4.677%	0.255%	9.200%	5.072%	0.092%	0.096%
70th	11.018%	49.577%	4.614%	21.450%	0.680%	6.755%	0.343%	1.564%	3.859%	0.076%	0.063%
90th	9.849%	38.607%	8.702%	19.792%	1.304%	3.677%	0.625%	13.917%	3.341%	0.147%	0.037%
96th	13.704%	48.161%	5.307%	20.734%	0.942%	4.604%	0.804%	2.027%	3.604%	0.090%	0.022%
89th	7.191%	54.900%	7.660%	15.911%	1.182%	5.945%	1.028%	3.544%	2.508%	0.119%	0.011%
81st	7.219%	42.573%	8.875%	18.346%	1.584%	6.200%	0.308%	12.434%	2.220%	0.193%	0.047%
88th	9.795%	50.478%	4.963%	17.454%	0.715%	4.780%	0.421%	6.886%	4.390%	0.088%	0.031%
99th	13.750%	45.112%	5.343%	17.415%	0.764%	4.562%	0.303%	7.460%	5.197%	0.091%	0.004%
77th	14.285%	29.237%	3.068%	19.433%	0.598%	3.233%	0.242%	7.227%	22.525%	0.118%	0.033%
94th	16.842%	37.993%	5.966%	25.841%	0.740%	4.739%	0.390%	2.802%	4.545%	0.093%	0.048%

NOTE: Chi Square Test for similarities conducted on Raw Data, not the Percentages

Pearson's chi-square test without Yates' continuity correction: X-square = 12788076, df = 100, p-value = 0

Table 4.2: Chi Square Testing of Lifestyle Segmentation Grouping Aspects of RSCs

[The percentage of the population lifestyle segment groupings for each RSC. This table demonstrates the difference in lifestyle segment grouping composition of each RSC.]

Similarly for the MV50 Segmentation information, the Chi Square Test reveals the segmentation distribution of the RSCs differs. The MV50 Segment Groups table also demonstrates those differences. Segment Groups MV50GP01, MV50GP02, MV50GP04, MV50GP08, and MV50GP09 are very different than other segments. Using percentages demonstrates the differences better than the raw data. There is one other noted feature of the data.

Figure 4.2 captures the essence of the original segmentation information for the contract data. Recruiter segment misclassification rate is 4.29% (segment 0 [4.06%] and segment 99 [0.23%]). Of the MV50 Lifestyle Segments, nearly 50% of USAR contracts come from the top ten segments. By concentrating on these top ten segments, recruiters

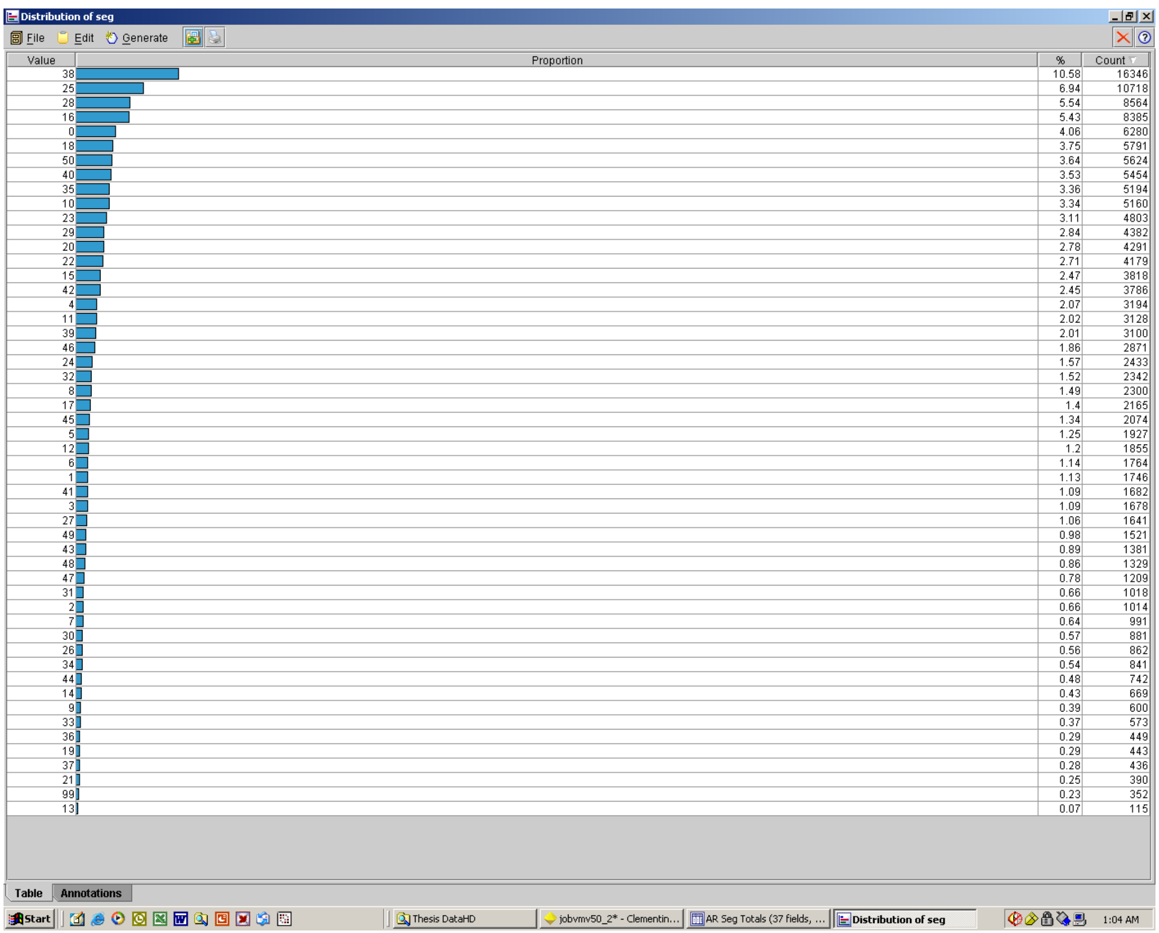


Figure 4.2: MV50 Lifestyle Segmentation Distribution of USAR Contract Data

can realize nearly half of the total contract effort for the USAR. This information could be incorporated into USAREC's mission distribution model or recruiting policy. Knowing the composition of the recruiting market could greatly assist USAREC, USARC, and the recruiting force accomplish its annual accession mission for the USAR.

Market composition, and the number of contracts obtained by each market, is a key component to understanding the recruiting environment. The more information acquired about the recruiting environment, the better we can make use of the personnel and monetary resources we have available. The additional information will enable us to formulate better predictive models to assist in the recruiting effort.

Knowing the market and RSC composition should assist in the type of units placed in the RSC's market. Predictive modeling will assist in unit stationing actions and prevent their poor placement in the market. The combination of these two pieces of information may greatly assist in future unit placement and stationing actions based on vocational, lifestyle segmentation, and unemployment aspects of RSC markets.

C. MODEL FITTING – A LEARNED PROCESS

Model fitting is a science and an art. After data familiarization, our intent was to treat all ZIP Codes equally. The way to achieve this was to place our tabular information into proportions so we could make comparisons with ZIP Code information. One bit of information necessary to review prior to starting our model fitting was to look at the unemployment rates of the country. How does the unemployment rate affect the outcome of contracts?

Unemployment data will change over time. Times series model development may be able to capture the unemployment rate over time, but we notice that the number of contracts produced annually per ZIP Code is generally small.

Figure 4.3, provided by the BLS, shows the national unemployment average for the period March 2003 through February 2004. Note that 5.9% is the national average. The map indicates there are counties in the US employing more than 94.1% of their population. Each county is clearly different.

The map demonstrates that the Midwest has the highest employment rates. This may be misleading since a greater portion of the Midwest land is used for farming. Since population density and number of jobs available are different than the rest of the country, this information may contain employment bias. This bias may be in the farming, forestry, and fishing vocation of the market.

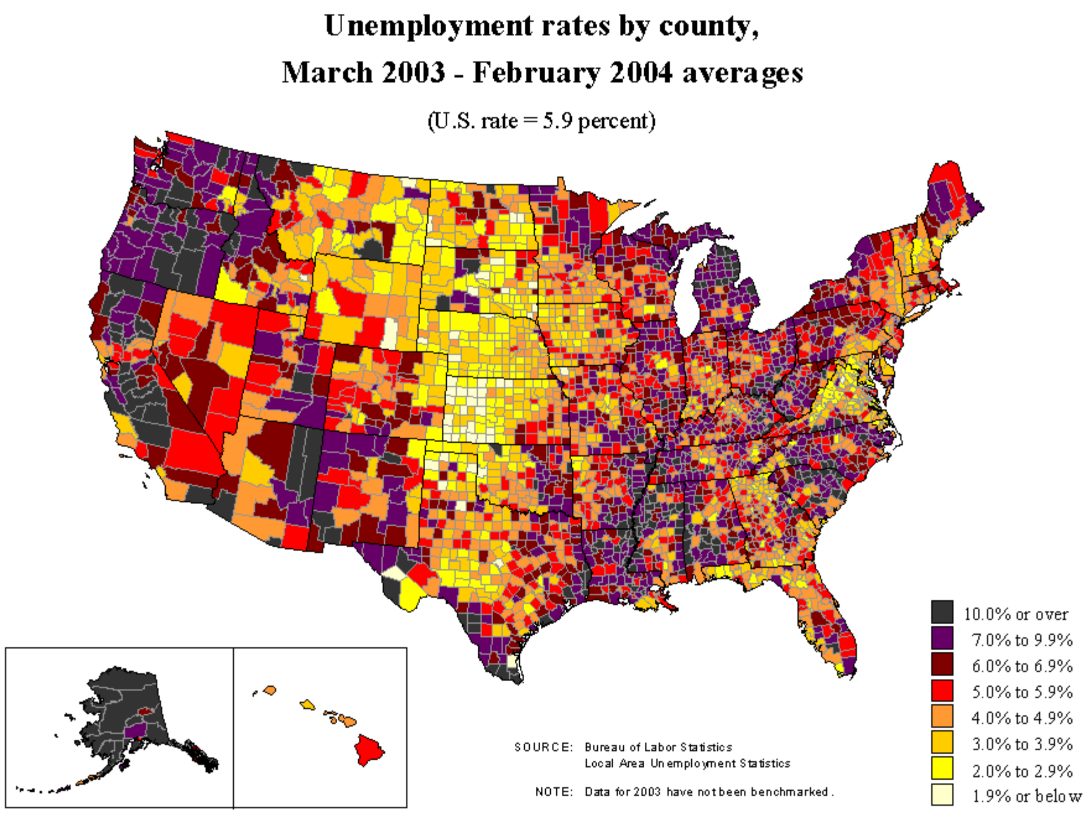


Figure 4.3: BLS Average Unemployment Rate by US County (Mar'03-Feb'04)

Our original approach was to treat each ZIP Code equally. We developed a model using all our predictors. The expected proportion contracts from a ZIP Code should depend on the demographic composition of the market.

In this case, we used the MA population (1), unemployment rate (1), vocational composition (11), and lifestyle segmentation composition (11) of the ZIP Code. This is a total of 24 (multiple regression) predictor variables to determine the outcome of the numbers of contracts a ZIP Code produces.

We tried four classes of models: Modeling the proportion of MA Population that enlisted, modeling the log (proportion of contracts), modeling total contracts as a Poisson random variable, and modeling total contracts as a Normal random variable. The preliminary model results are:

- The model of proportions had low explanatory power – only an R-Squared of 16%.
- The log-normal model required discarding about 10% of the data having zero contracts. The resulting R-Squared was smaller – only a little over 12%.
- The Poisson model explained about a little over 21% of the variation of the data.
- The Normal model did better; and we fully developed it.

D. MODEL FITTING – AVERAGE ANNUAL CONTRACTS

Having described lifestyle segments and vocations, we can formulate and continue to evaluate our regression models. The next model evaluated is a simple linear regression model. The expected number of contracts from a ZIP Code should depend on the demographic composition of the market.

In this case, we used the MA population (1), unemployment rate (1), vocational composition (11), and lifestyle segmentation composition (11) of the ZIP Code. This is a total of 24 (multiple regression) predictor variables to determine the outcome of the numbers of contracts a ZIP Code produces. The model we develop has a slope, an intercept value, and regression coefficients for each predictor variable. Recall that a multiple linear regression model has the following form:

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_j x_j$$

Equation 4.2: General Form of Multiple Linear Regression Model

In our case, we have *j* equal to 24. We express the expected number of contracts as a linear combination of the ZIP Code predictor variables. Keep in mind that there are 29,865 ZIP Codes in our table. This information has the following linear model (LM) construct. The number of contracts is a linear function of (MA.POP, un.rate, EXECMNGE, FAFOFISH, ADMINSPT, PROFSNL, TECHSPT, SVCOTHR, SVCPROT, SALES, CRFTSMAN, LABORERS, TRANSPO, MV50GP01, MV50GP02, MV50GP03, MV50GP04, MV50GP05, MV50GP06, MV50GP07, MV50GP08, MV50GP09, MV50GP10, MV50GP11). Table 4.3 contains the detailed results from the regression.

Not all variables in the regression appear to be significant. With this LM, we achieve a multiple R-Squared of 0.6934, compared with a 0.7737 correlation of MA population with contracts (i.e. MA population alone explains over 59% of the variation). About 10% is explained by demographics and vocations. The rest of the variation is likely to be due to policy (numbers of recruiters, station and recruiter placement, mission emphasis, goals, etc.).

We see that SALES, TRANSPO, MV50GP01, and MV50GP11 appear to be insignificant in our table, as they all have p-values that exceed 0.05. This indicates that their respective coefficient values in the regression equation may be 0. We remove them from the regression and see that the R-Squared does not change much.

We next look at the model's coefficients. They tend to be small due to the scale of the predictors. We are predicting the average annual number of USAR contracts achieved in each ZIP Code. Does the order of magnitude make sense? The answer is yes. We have 29,865 ZIP Codes and a USAR NPS mission of 20,000. If each ZIP Code produces an average of one contract per year, we would have 29,865 contracts.

*** Linear Model ***

```
Call: lm(formula = AR.Avg ~ MA.POP + un.rate + EXECMNGE + FAFOFISH
+ ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN +
LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 +
MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 +
MV50GP11, data = ALLDATAbyZIP2, na.action = na.exclude)

Residuals:     Min     1Q      Median   3Q     Max
             -7.229 -0.2106 -0.05927 0.1448 23.3
```

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.1388	0.0151	9.2169	0.0000
MA.POP	0.0001	0.0000	18.5959	0.0000
un.rate	-1.5446	0.2260	-6.8339	0.0000
EXECMNGE	-0.0002	0.0000	-18.5070	0.0000
FAFOFISH	-0.0006	0.0000	-13.4253	0.0000
ADMINSPT	0.0007	0.0000	23.1085	0.0000
PROFSNL	0.0001	0.0000	4.7454	0.0000
TECHSPT	0.0008	0.0000	24.3688	0.0000
SVCOTHR	0.0001	0.0000	8.7381	0.0000
SVCPROT	0.0003	0.0000	7.6545	0.0000
SALES	0.0000	0.0000	0.3048	0.7606
CRFTSMAN	-0.0002	0.0000	-16.6801	0.0000
LABORERS	-0.0021	0.0003	-7.7198	0.0000
TRANSPO	0.0000	0.0000	1.4508	0.1468
MV50GP01	0.0000	0.0000	1.4467	0.1480
MV50GP02	0.0000	0.0000	5.2424	0.0000
MV50GP03	0.0015	0.0000	40.9519	0.0000
MV50GP04	0.0000	0.0000	-8.0032	0.0000
MV50GP05	-0.0014	0.0002	-6.5840	0.0000
MV50GP06	-0.0003	0.0000	-14.9495	0.0000
MV50GP07	-0.0012	0.0003	-4.3682	0.0000
MV50GP08	0.0001	0.0000	14.9848	0.0000
MV50GP09	-0.0001	0.0000	-9.2602	0.0000
MV50GP10	-0.0014	0.0005	-2.5787	0.0099
MV50GP11	0.0001	0.0001	0.6916	0.4892

```
Residual standard error: 0.8929 on 29839 degrees of freedom
Multiple R-Squared: 0.6934
F-statistic: 2811 on 24 and 29839 degrees of freedom, the p-value is 0
1 observations deleted due to missing values
```

Table 4.3: S-Plus Linear Regression Model Formulation for Number of USAR Contracts

Table 4.4 shows the results of removing variables, the resulting multiple R-Squared, and the regression df. One of the last predictor variables removed is MA.POP. We see that even removing MA.POP as a predictor does not change the amount of explained variation. We removed 11 predictor variables with very little change in the amount of explained variation in our LM. This suggests those variables are insignificant and do not contribute to the overall explanation of variation in the number of contracts a ZIP Code produces. A simpler model yielding the same R- Squared is usually preferred.

VARIABLE REMOVED	MUTLIPLE R-SQUARED	RGRSN DF
SALES	0.6934	23
MV50GP11	0.6934	22
TRANSPO	0.6933	21
MV50GP01, MV50GP10	0.6933	19
MV50GP05, MV50GP06	0.6902	17
MV50GP07, MV50GP09	0.6877	15
SVCOTHR, MA.POP	0.6821	13

Table 4.4: Predictor Variable Removal and Multiple R-Squared Results

After subsetting, we obtain the model in Table 4.5. Note that, as in Table 4.3, the coefficients of some of the predictor variables are negative. This indicates the number of

```
Residuals:     Min    1Q   Median    3Q    Max
            -40.77 -1.329  -0.29  0.8917 144.6

Coefficients  Value  Std.Error   t-value  Pr(>|t|)
(Intercept)   0.1258  0.0150     8.4104   0.0000
    un.rate  -1.8258  0.2269    -8.0468   0.0000
   EXECMNGE  -0.0002  0.0000   -33.5045   0.0000
   FAFOFISH  -0.0004  0.0000   -10.0658   0.0000
   ADMINSPT   0.0009  0.0000    44.8462   0.0000
    PROFSNL   0.0002  0.0000    17.9365   0.0000
    TECHSPT   0.0007  0.0000    21.0479   0.0000
    SVCPROT   0.0003  0.0000     7.2537   0.0000
   CRFTSMAN  -0.0001  0.0000    -7.3236   0.0000
   LABORERS  -0.0028  0.0003   -11.0822   0.0000
   MV50GP02   0.0000  0.0000     3.7429   0.0002
   MV50GP03   0.0013  0.0000    47.4132   0.0000
   MV50GP04   0.0001  0.0000    -4.8638   0.0000
   MV50GP08   0.0002  0.0000    26.6054   0.0000

Residual standard error: 5.454 on 29850 degrees of freedom
Multiple R-Squared: 0.6821
F-statistic: 4926 on 13 and 29850 degrees of freedom, the p-value is 0
1 observations deleted due to missing values
```

Table 4.5: S-Plus Linear Regression Model Formulation for Number of USAR Contracts (iteration 12)

contracts a ZIP Code produces is negatively associated with the size of the variable in the ZIP Code. For example, we see that un.rate, EXECMNGE, FAFOFISH, LABORERS, and CRAFTSMAN all have negative coefficients. The larger the unemployment rate, or the greater the proportion in these vocations, the fewer contracts the ZIP Code can

produce. In particular, for every 10,000 LABORERS in the ZIP Code, the expected annual average number of USAR contracts decreases by 28.

Table 4.5 demonstrates the resulting regression equation after 11 iterations of variable removal. The amount of explained variation is still greater than 68%. Since the "full" model had over 69% explained variation and the amount of explained variation is greater than 68% with 11 of our original variables removed, we use the simpler model. Determining the significance of predictor variables is a way to achieve a simpler more effective model.

A good tool used to verify the model is to plot the data and look at its appearance. We can achieve this in two plots. The first is the actual data versus the "fitted" data. The fitted data is the predicted value or outcome using the regression equation. The second is the fitted data versus the residuals. The residuals are the deviation from the mean value of the regression. The mean value of the regression in a LM is the slope of the regression equation.

Figure 4.4 shows the graph of the USAR actual average annual number of contracts and the USAR fitted average annual number of contracts. There are some values in the data that largely deviate from the regression model. These values are outliers. If you remove them from the regression and the slope of the regression line greatly changes, then they are large influencers. Normally, a determination needs to be made on outlier exclusion or inclusion. Since we have nearly 30,000 data points in our regression, we will disregard these outliers.

The data should have a strong linear look to have a good LM fit. The graph appears generally linear. Notice the strong concentration of data points from 0 to approximately 6.5 annual contracts. This indicates that the predictions for the average annual number of USAR contracts should be fairly accurate in this region of the model.

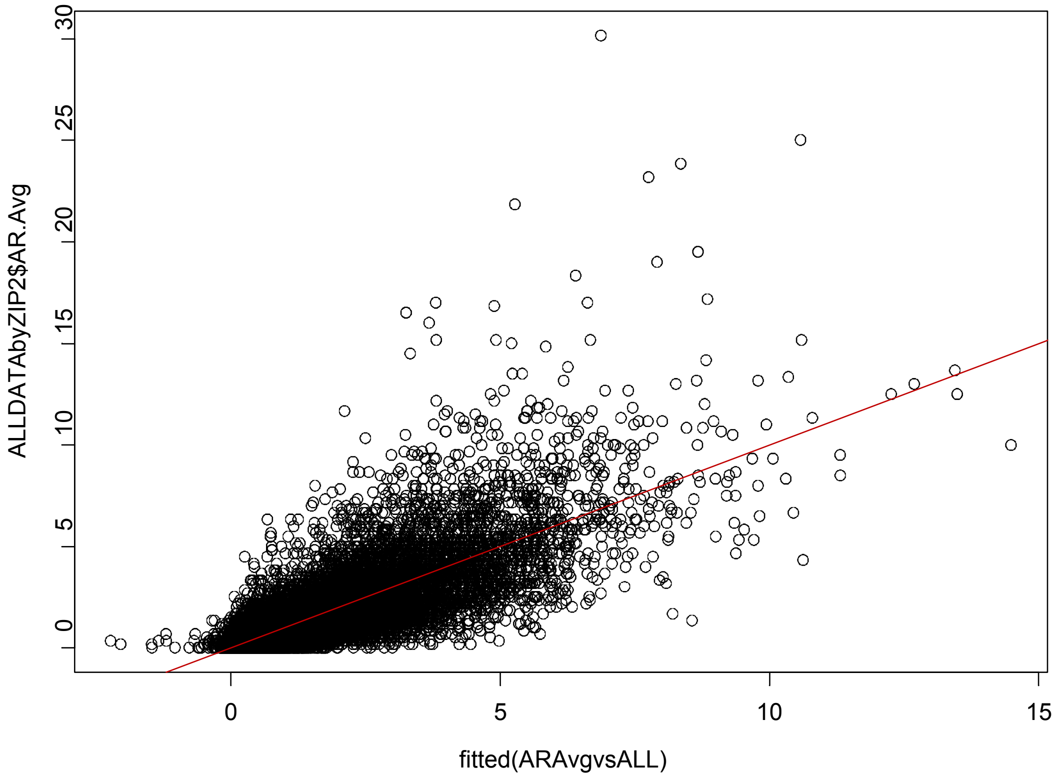


Figure 4.4: Graph of Army Reserve Average Annual Contracts versus Army Reserve Fitted Average Annual Contracts

Figure 4.5 shows the graph of the USAR fitted average annual number of contracts and residuals. The points on the plot should be randomly scattered throughout the plot for the model to have proper fit to the data. This indicates model departures. We see that there is a linear relationship at the bottom left of our plot. Normally this indicates some kind of dependence in the data. The assumption is independent variables with homogeneous variance.

Figure 4.5 would normally indicate heterogeneity of the variance, but we know this data. We tried and discarded a log model because the number of contracts for some ZIP Codes was zero. A log transformation would therefore not be appropriate here. We

might consider other transformations like log(n+1), where n is the number of contracts in the Zip Code, or the substitution of 0.001 for those ZIP Codes which produced zero contracts.

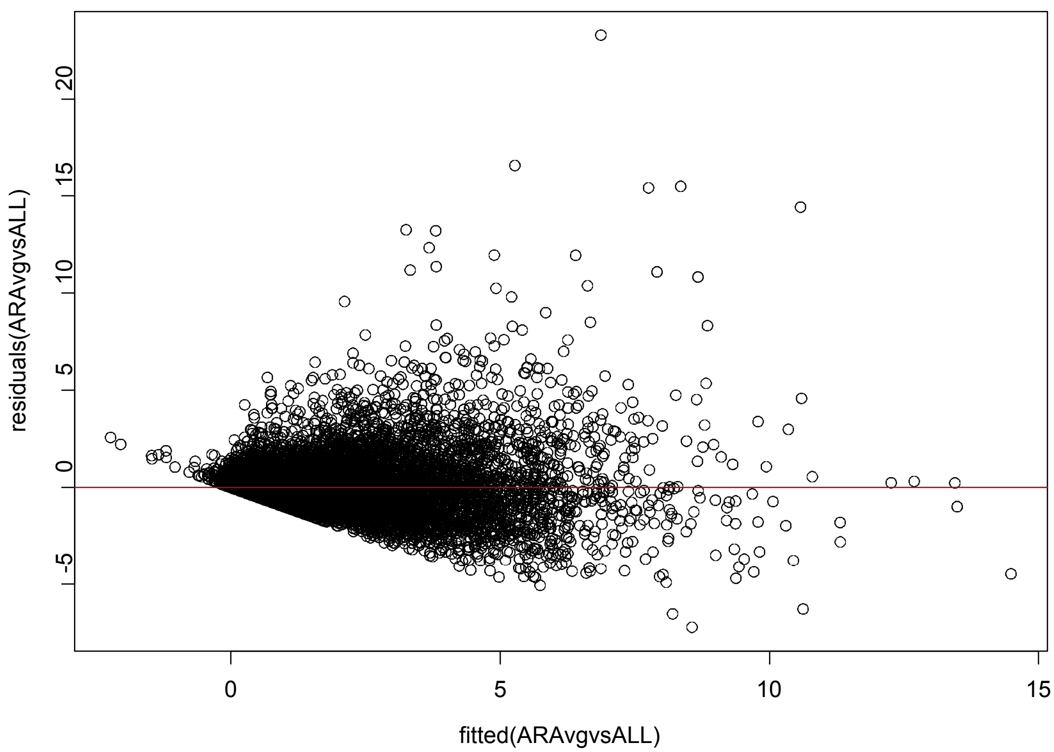


Figure 4.5: Graph of Army Reserve Fitted Average Annual Contracts versus Residuals

Figure 4.5 appears to indicate heteroscedasticity because the number of contracts is either zero or positive. Constant variance would plot the residuals scattered about the graph without pattern or shape. If it were not for this phenomenon, we would see the bottom left of the plot filled with data points as well.

Let's look at an example problem for a few ZIP Codes to see how our regression equation performs. Since we are here at the Naval Post-Graduate School in Monterey, CA, we will use ZIP Code 93940. Keep in mind we are using our smaller derived model. The unemployment rate for Monterey ZIP Code 93940 is 10.44%. Table 4.6 has the

remaining values for the predictor variables. The table construct is such that we can compute the dot product of the values for and the coefficients of the regression equation to produce the estimated number of contracts for the ZIP Code.

MONTEREY: 93940

Predictor	Coefficient	Values
(Intercept)	0.1258	1
Un.rate	-1.8258	0.1044
EXECMNGE	-0.0002	12280
FAFOFISH	-0.0004	191
ADMINSPT	0.0009	2698
PROFSNL	0.0002	8756
TECHSPT	0.0007	1326
SVCPROT	0.0003	596
CRFTSMAN	-0.0001	2806
LABORERS	-0.0028	143
MV50GP02	0.0000	1584
MV50GP03	0.0013	154
MV50GP04	-0.0000	6347
MV50GP08	0.0001	6
	PREDICTED:	**1.47**
	ACTUAL:	***1.00***

Table 4.6: Annual USAR Contract Prediction Results for Monterey, CA 93940

Looking at the historical information of the ZIP Code for Monterey we find the range of contracts is (0, 3). The six-year average for the ZIP Code is 1 contract per year. This is another reason not to do monthly time series analysis – we would have mostly zeros in your data. The annual predicted number of contracts is 1.47. The difference is 0.47 contracts. The 95% confidence interval of the prediction is (1.41, 1.54) with a standard error of 0.03. We could obtain a confidence interval for our raw contract data, if we tested the values for normality and tested the residuals. Since we only have 6 data points, annual number of contracts, in our sample for each ZIP Code, this approach would be futile. This makes the regression worth the effort.

Table 4.7 demonstrates the same information for one Salinas, CA Zip Code, 93901 (note that some cities, like Salinas, have more than one ZIP Code associated with it). In the Salinas ZIP Code 93901 case, the annual predicted number of contracts is 1.68. The difference is 0.68 contracts. The 95% confidence interval of the prediction is (1.55, 1.81) with a standard error of 0.07.

SALINAS: 93901

Predictor	Coefficient	Values
(Intercept)	0.1258	1
un.rate	-1.8258	0.1044
EXECMNGE	-0.0002	6702
FAFOFISH	-0.0004	1434
ADMINSPT	0.0009	2405
PROFSNL	0.0002	4252
TECHSPT	0.0007	1074
SVCPROT	0.0003	1120
CRFTSMAN	-0.0001	3087
LABORERS	-0.0028	70
MV50GP02	0.0000	4159
MV50GP03	0.0013	269
MV50GP04	-0.0000	3152
MV50GP08	0.0002	476
	PREDICTED:	**1.68**
	ACTUAL:	***1.00***

Table 4.7: Annual USAR Contract Prediction Results for Salinas, CA 93901

SEASIDE: 93955

Predictor	Coefficient	Values
(Intercept)	0.1258	1
un.rate	-1.8258	0.1044
EXECMNGE	-0.0002	5274
FAFOFISH	-0.0004	478
ADMINSPT	0.0009	2280
PROFSNL	0.0002	3455
TECHSPT	0.0007	705
SVCPROT	0.0003	520
CRFTSMAN	-0.0001	3059
LABORERS	-0.0028	63
MV50GP02	0.0000	4353
MV50GP03	0.0013	454
MV50GP04	-0.0000	1515
MV50GP08	0.0002	954
	PREDICTED:	**2.15**
	ACTUAL:	***2.83***

Table 4.8: Annual USAR Contract Prediction Results for Seaside, CA 93955

Likewise, looking at the historical information of the 93955 ZIP Code for Salinas we find the range of contracts is (0, 2) and the six-year average for the ZIP Code is 1 contract.

Finally, we look at Seaside, CA. Table 4.8 demonstrates the information for the Seaside, CA Zip Code 93955. In the Seaside ZIP Code 93955 case, the annual predicted number of contracts is 2.15. The 95% confidence interval of the prediction is (2.11, 2.20) with a standard error of 0.02. The range of contracts is (2, 6) and the difference is 0.68 contracts.

These predicted values will become the $max\_recruits\_zip_i$ for the eventual LP model. So we have the parameter for the maximum number of recruits obtained at ZIP Code i is the predicted value of the number of recruits obtained from ZIP Code i. The equation is as follows:

$$max\_recruits\_zip_i = \max(0, \widehat{AR.Avg_i})$$

Equation 4.3: Maximum Number of Recruits Formula

There were about 120 negative predicted values, and we set them to zero in Equation 4.3.

E. MODEL FITTING – TOP FIVE MOSs

The next item brought out in this analysis is the maximum number of recruits at ZIP Code i of MOS j. This is the maximum of zero or the minimum of the predicted number of contracts of MOS j in ZIP Code i and the predicted number of recruits obtained at ZIP Code i. The formulation of the equation for this parameter in the eventual LP is:

$$max\_recruits\_zip\_mos_{i,j} = \max(0,\ \min(\widehat{MOS_{i,j}},\ \widehat{AR.Avg_i})$$

Equation 4.4: Maximum Number of Recruits by MOS Formula

This keeps MOS predictions non-negative and within the total production.

We now turn our attention to modeling the top five MOSs. This information is located in Appendix G (Top Five MOS Regression Equations). The current top five

MOSs are 52D, 74D, 77F, 88M, and 95B. We followed the same procedures for the MOS predictions as we did for the annual number of USAR contracts.

However, since we do not know the importance of certain predictor variables in our models and since including insignificant variables will not change the outcome of the prediction; we will employ the full model for our top five MOSs. Recall that Phase II will construct all 264 MOSs in detail. Phase II will make the determination of the significance of predictor variables.

The full model has the following LM construct. The actual number of contracts

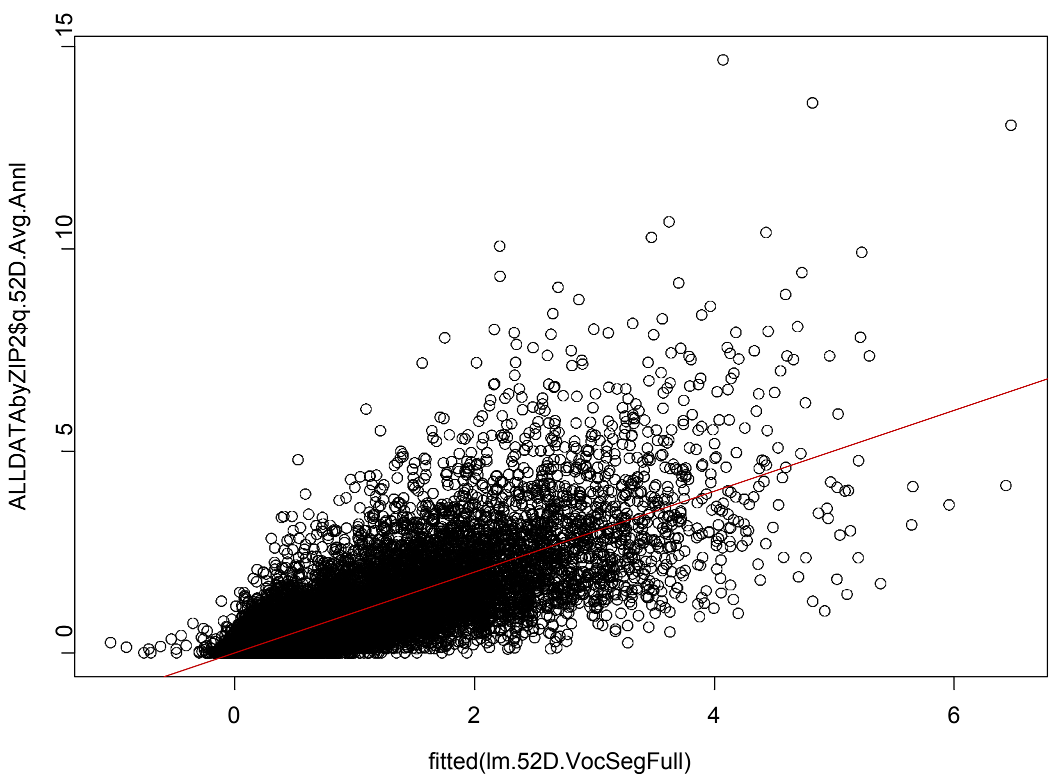


Figure 4.6: Graph of Average Annual USAR Contracts Qualifying for MOS 52D versus Fitted Average Annual USAR Contracts Qualifying for MOS 52D

that qualified for MOS *j* in ZIP Code *i*, regardless of contracted MOS, is a linear function of (MA.POP, un.rate, EXECMNGE, FAFOFISH, ADMINSPT, PROFSNL,

TECHSPT, SVCOTHR, SVCPROT, SALES, CRFTSMAN, LABORERS, TRANSPO, MV50GP01, MV50GP02, MV50GP03, MV50GP04, MV50GP05, MV50GP06, MV50GP07, MV50GP08, MV50GP09, MV50GP10, MV50GP11). Appendix G contains the detailed results from the regression.

As with the predicted number of contracts, we ran diagnostic plots on actual versus fitted and fitted versus residuals. Figures 4.6 and 4.7 plots appear to be satisfactory. Notice Figure 4.7 has the same "shoulder" on the fitted versus residual plot.

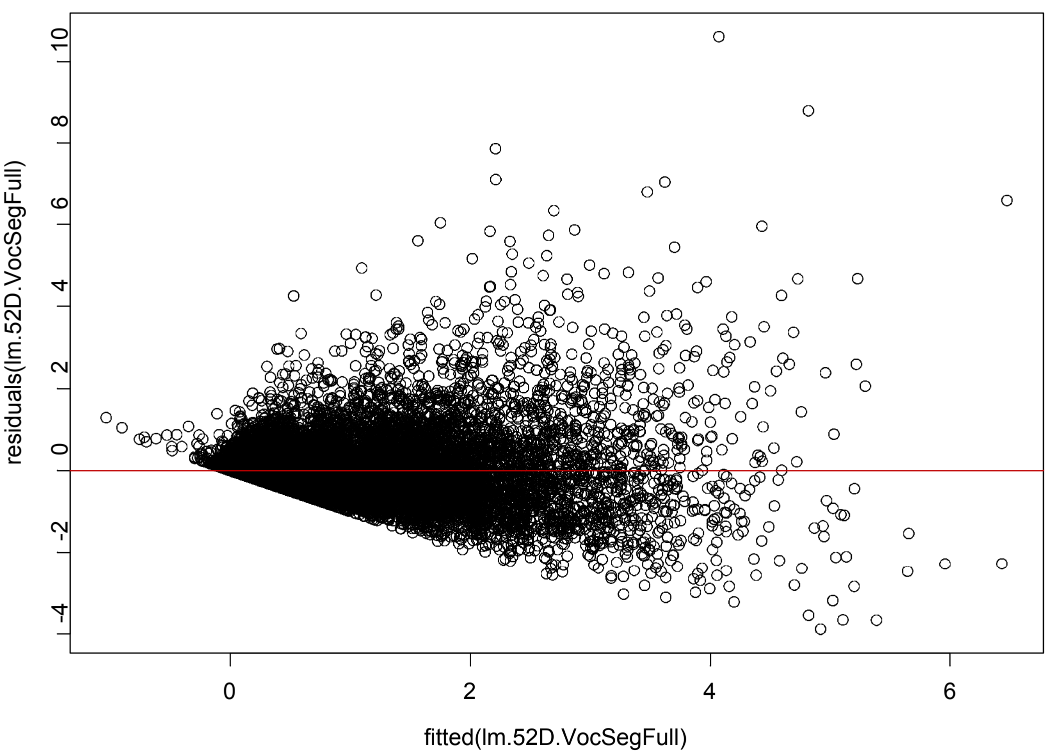


Figure 4.7: Graph of Fitted Average Annual USAR Contracts Qualifying for MOS 52D versus Residuals

Again this is because of the positive nature of contracts and those whom qualify for a particular MOS in a ZIP Code.

The other top four MOS (74D, 77F, 88M, and 95B) plots, located in Appendix G, are very similar for both fitted versus actual and fitted versus residuals. As with our

predicted number of contracts model, let's look at an example. To keep it simple, we will use the same ZIP Codes (93901, 93940, and 93955) as previously.

How does our regression equation perform? Keep in mind we are using the full model because we will eventually want to examine MOS *j* in ZIP Code *i*. To accomplish the comparison, we need to examine the same model for each MOS. In Phase II each MOS will have its own model. We construct Table 4.9 in the same manner as before

MONTEREY: 93940

Predictor	Coefficient	Values
(Intercept)	0.1070	1
MA.POP	0.0001	5484
un.rate	-1.2145	0.1044
EXECMNGE	-0.0001	12280
FAFOFISH	-0.0003	191
ADMINSPT	0.0002	2698
PROFSNL	0.0001	8756
TECHSPT	0.0004	1326
SVCOTHR	0.0000	5018
SVCPROT	-0.0001	596
SALES	0.0001	5650
CRFTSMAN	-0.0001	2806
LABORERS	-0.0009	143
TRANSPO	0.0000	2104
MV50GP01	0.0000	3269
MV50GP02	0.0001	1584
MV50GP03	0.0007	154
MV50GP04	-0.0001	6347
MV50GP05	-0.0009	15
MV50GP06	-0.0001	446
MV50GP07	0.0001	0
MV50GP08	-0.0001	6
MV50GP09	-0.0001	128
MV50GP10	-0.0013	5
MV50GP11	0.0000	0
	PREDICTED:	**1.13**
	ACTUAL:	***0.39***

Table 4.9: Average Annual USAR Contracts Qualified for MOS 52D Prediction Results for Monterey, CA 93940

such that we can achieve the dot product of the values for and coefficients of the regression equation for the annual average number of USAR contracts qualifying for MOS 52D in the ZIP Code.

Looking at the information of ZIP Code 93940 for Monterey we find the actual average number of contracts qualifying for MOS 52D is 0.39 contracts. According to our

model formulation, this value is not a rate, but rather the maximum number of recruits qualifying in ZIP Code 93940. The annual predicted number of contracts is 1.13 and the 95% confidence level interval is (1.09, 1.17) with a standard error of 0.02. The difference is 0.72 contracts.

Similarly Tables 4.10 and 4.11 demonstrate the same information for one Salinas,

SALINAS: 93901

Predictor	Coefficient	Values
(Intercept)	0.1070	1
MA.POP	0.0001	4374
un.rate	-1.2145	0.1044
EXECMNGE	-0.0001	6702
FAFOFISH	-0.0003	1434
ADMINSPT	0.0002	2405
PROFSNL	0.0001	4252
TECHSPT	0.0004	1074
SVCOTHR	0.0000	3678
SVCPROT	-0.0001	1120
SALES	0.0001	4879
CRFTSMAN	-0.0001	3087
LABORERS	-0.0009	70
TRANSPO	0.0000	3994
MV50GP01	0.0000	821
MV50GP02	0.0001	4159
MV50GP03	0.0007	269
MV50GP04	-0.0001	3152
MV50GP05	-0.0009	39
MV50GP06	-0.0001	667
MV50GP07	0.0001	6
MV50GP08	-0.0001	476
MV50GP09	-0.0001	263
MV50GP10	-0.0013	11
MV50GP11	0.0000	0
	PREDICTED:	**0.79**
	ACTUAL:	***0.63***

Table 4.10: Average Annual USAR Contracts Qualified for MOS 52D Prediction Results for Salinas, CA 93901

CA Zip Code, 93901 and one Seaside, CA Zip Code, 93955. The differences are 0.16 and 0.13, respectively. The annual predicted number of contracts for Zip Code 93901 is 0.79 and the 95% confidence level interval is (0.71, 0.88) with a standard error of 0.04. The annual predicted number of contracts for Zip Code 93955 is 1.56 and the 95% confidence level interval is (1.47, 1.64) with a standard error of 0.04.

Vocational and demographic composition have a considerable effect on the outcome of the regression. Recall our regression equation for each MOS uses the full model. Any large increase or decrease in demographic composition will have an effect

SEASIDE: 93955

Predictor	Coefficient	Values
(Intercept)	0.1070	1
MA.POP	0.0001	6528
un.rate	-1.2145	0.1044
EXECMNGE	-0.0001	5274
FAFOFISH	-0.0003	478
ADMINSPT	0.0002	2280
PROFSNL	0.0001	3455
TECHSPT	0.0004	705
SVCOTHR	0.0000	8820
SVCPROT	-0.0001	520
SALES	0.0001	4464
CRFTSMAN	-0.0001	3059
LABORERS	-0.0009	63
TRANSPO	0.0000	3440
MV50GP01	0.0000	596
MV50GP02	0.0001	4353
MV50GP03	0.0007	454
MV50GP04	-0.0001	1515
MV50GP05	-0.0009	58
MV50GP06	-0.0001	236
MV50GP07	0.0001	46
MV50GP08	-0.0001	954
MV50GP09	-0.0001	267
MV50GP10	-0.0013	5
MV50GP11	0.0000	0
	PREDICTED:	**1.56**
	ACTUAL:	***1.43***

Table 4.11: Average Annual USAR Contracts Qualified for MOS 52D Prediction Results for Seaside, CA 93955

on the prediction. As we review Appendix G and peruse the outcome of the vocations, segments, MA population, and unemployment rate coefficients, we note that MOSs have different coefficients indicating larger or smaller influences of these factors in the ZIP Code.

For example, if we compare MOS 52D with MOS 95B we notice MV50 Segment Groups 1, 7, and 11 appear to be statistically insignificant for MOS 52D. By contrast, notice MV50 Segment Groups 1, 8, and 11 appear to be statistically insignificant for

MOS 95B. This also occurs with the vocations. The LM for MOS 52D does not appear to contain the SVCOTHR vocation while the LM for MOS 95B does not appear to contain the TRANSPO vocation.

Table 4.12 has the $max\_recruits\_zip\_mos_{i,j} = \max(0, \min(\widehat{MOS}_{i,j}, \widehat{AR.Avg}_i)$ for each of the MOSs for the three example ZIP Codes.

ZIP Code	52D	74D	77F	88M	95B	MAX
93901	0.790	0.940	1.220	1.250	1.150	***1.250***
93940	1.130	1.260	1.460	1.470	1.420	***1.470***
93955	1.560	1.840	2.150	2.150	2.150	***2.150***

Table 4.12: Maximum Number of Recruits Qualifying for the USAR Top Five MOSs for ZIP Codes 93901, 93940, and 93955

Note that the number qualifying varies by ZIP Code. The point of the analysis is that not all MOSs are equally supportable. We must consider ZIP Code supportability by MOS to obtain correct unit positioning. This variation is the reason for USAR unit force structure optimal stationing LM that we are developing.

There are similarities in the LM development for the MOS *j* in ZIP Code *i*, but Phase II analysis must develop a model for each MOS. The basis for the LM formulation in Phase II can be the full model developed herein for the top five MOSs.

We now have the two inputs for the Phase III model. Phase II of this analysis will develop the regression equations for the remaining 259 MOSs.

F. MODELING OUTCOME

Data analysis can only reveal some of the predictive tendencies of a modeled environment. Some analyses may not be able to show peculiarities in the data. We thought and supposed that segmentation, vocational, and unemployment information of markets at ZIP Code detail would have predictive capability on the number of contracts a ZIP Code can produce.

Our developed LM explains about 70% of the variation in the data. The remaining variation, with respect to our variables in the data, is assumed random. However, it appears as though there is some other phenomenon that would explain the remaining data variation. As previously stated, the RSCs differ in the segment and vocational composition. This information suggests that there are remaining non-demographic factors influencing the number of contracts produced by ZIP Code. Regionalization may have a discernible affect on the data. FIP, MSA, State, ASG, RSC, etc. may be a way to gain more predictive power with model.

The focus of this analysis was to be able to predict the number of contracts a ZIP Code could produce based on market segments, vocational information, and unemployment rates. We developed a useful model that has 70% predictive power; that is, we were able to explain about 70% of the variation of the data.

The number of NPS contracts a ZIP Code can produce may depend on additional aspects of the recruiting process not considered. Production may rely on mission quotas, mission levels, policy, etc. Another model to produce NPS contracts may be found in the historical contract data. One may be able to examine the historical data with respect to each market, provided structure has not changed, and develop a predictive model based on some kind of mean or moving average to smooth the data.

One of the data exploratory methods not considered is this analysis is time series. Time series requires the collection of data elements by time interval. We could have arranged the contract data by month. To accomplish this, we could have included the actual contract date of the accession. The vocational information should not change much over time. Likewise, the segmentation would not change much over time. We could collect the unemployment rates by month for the same time period.

Constructing the collection and subsequent analysis in this manner may lead to a better predictive model. This data may have seasonality associated with it. When we manipulated and assembled the current data, we used a single data point to summarize six years of data for each ZIP Code. This information may be bound by the construct of one data point to represent the entire 6-year period (FY98-FY03). What may be more appropriate is to obtain the monthly data for the ZIP Code and investigate time series

performance of the data. This approach may lead to a more lucrative predictor and a better understanding of the data peculiarities.

Regionalization may have a discernible affect on the data. State, RSC, MSA, etc. may be a way to gain more predictive power with modeling. We see that the vocations and lifestyle segment for each RSC are statistically different in composition. The market composition has a great deal to do with the number of contracts USAREC obtains from the markets. We see that our modeling efforts has predictive power, gives explanation, and understanding as to what variables yield inclusion into modeling the annual number of contracts.

We may be able to use an indicator variable for each of the regions (1,2,…, 10) to capture regional effects. This regional effect would then be translated into the intercept of our regression.

The development of the MOS data also has predictive capability. We are able to explain about 65% of the variation in the data by the top five MOSs. This suggests that our model is useful in explaining the variation of the data by MOS and ZIP Code. We note that the prediction of the number of contracts a ZIP Code yields may also vary more because of the amount of effort a recruiter places on achieving his mission.

It appears as though our developed model is plausible and generates new conclusions about the data. The next section addresses further considerations.

G. POSITIONING UNITS TO OPTIMIZE OTHER METRICS, GIVEN 95% (OR OTHER LEVEL) OPTIMAL FILL.

1. Cost (Incentives, reorganization, transportation, etc.).

The cost implications for relocating structure are a function of whether there is an RC in the new determined location. If a RC exists at a location, the cost would be the cost of relocating structure to the new area plus the cost of relocating current structure, if applicable, to another area.

Costing of this information can be ascertained in the overall model developed in Phase III of this project. These metrics can be included in the LP. Once obtained, they can be optimized in the same manner.

2. Geographical balance (the HLS connection).

After consideration of moving and repositioning unit structure, we need to redress the geographical balance of our force structure distribution. As previously stated, the USAR constructs its RSC around the FEMAs. If the new structure has the desired vocations necessary to complete its FEMA missions, then we do not have an impact. We maintain the geographical balance. However, if the new structure does not have the desired vocations, then we may consider moving the needed structure to support the FEMAs, change some of the FEMA missions to accommodate the new structure, or do a combination of both.

H. POSITIONING UNITS TO OPTIMIZE FILL RATE, GIVEN OTHER METRICS AS CONSTRAINTS

The question remains, how to position the structure with respect to the market? Since we obtained the regression equations for the top five MOSs in the inventory, we can begin to position the force structure within the markets by using the equations and the follow-on Phase II equations as predictors. The factors of volume, quality, unemployment rate, vocations, etc. determine the supportable force structure composition. The ability of the market to support TPU structure at its current location is key to successfully determining the structure location.

The regression equations for each MOS forecast the support of the MOS in the market. We augment those markets not obtaining appropriate level of MOSs with advertising or regionally based incentives. Offering educational, MOS bonus, or some other enticement may cause sufficient quantities of qualified MA to join those units. We are not too far off desired unit fill rates, in most cases. We may increase our fill rates by offering these inducements.

There may several reasons why some areas are successful and other areas are not. However, this study demonstrates the quality assessment information. Along with historical quality information, the vocational information of the area, and production information will tell what type structure is most successful by RC. It also demonstrates that the most prominent vocations of the area can be related to the type structure placed by the USAR to be supported in the area.

If other Sister Services are willing to give up their production data, we could determine the "overall" affect/effect of the study on Department of Defense recruiting, retention, and structure placement efforts.

I. CRITICAL ASSUMPTIONS

The first critical assumption is human factors do not influence the outcome of the analysis (i.e. "All recruiters and commanders are created equal"). The second is that the "best" distribution methodology for force structure is independent of the requirements on recruiting and the needs of the force structure composition (i.e. recruiting effort and force structure requirements are independent).

Thirdly, we assumed that there is no bias in the structure lay-down, quality assessment, positioning of recruiting assets, individual efforts of each recruiter, and production historical information.

Lastly, it is reasonable to assume that vocations, lifestyle segmentation, LSCATs, etc. are market influencers. Without the knowledge of these items, we could not obtain necessary information about our population.

J. SUMMARY

The analysis demonstrates the assessment of the unit positioning and market quality has pay-offs. The results of this analysis need to be further studied and included as part of the constraint set in an optimizing distribution model. This provides the basis for the improvement of stationing and recruiting for America's Army.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

A. SUMMARY

As with all analyses, we began this analysis with a problem. The problem was the unit fill rate environment of the USAR. Procedurally we processed our analysis by defining a structure to assist in the process. We identified the problem, identified factors or components, developed a model, collected the data, and determined the model's validity.

This thesis is Phase I. Recall these three phases are:

Phase I: Process Definition, Data Collection, and Data Scrubbing.

Phase II: MOS Build – Populate Data Fields for the Optimization Model.

Phase III: Construct and Complete the Optimization Model.

We assembled the data on over 30,000 ZIP Codes, over 800 RCs, and over 260 Military Occupational Specialties (MOSs), drawing on and integrating over a dozen disparate data bases. This effort produced a single table with demographic, vocational, and economic data on every ZIP Code in America, along with the six-year results of RA, USAR, and Sister Service recruit production. Data was also obtained on the quality of each recruit and his suitability for each of the 264 Army MOSs.

We see regression, with the considered variables, yields a predictive model to forecast numbers of contracts with suitable qualifications for each MOS. Preliminary modeling developed a model that accounts for about 70% of the variation in recruit production by ZIP Code. We also obtain the demographic and vocational composition of the ZIP Codes.

Models for the top five USAR MOSs, contained in Appendix G, were also developed to predict the maximum number of recruits obtained from a ZIP Code for that MOS. ZIP Codes vary in their ability to provide recruits with sufficient aptitude for technical fields, and this is illustrated in this thesis with examples.

This modeling gives new explanatory and predictive capability. We had presumed that the unemployment rate of the ZIP Code would add explanation to the regression. In each of the models, the unemployment rates were statistically significant. However, it does not appear as though they are practically significant. In each case, we see a negative coefficient in the model. This is likely due to confounding effects among the predictors.

Remember, Phase I built only the top five MOSs. The Phase I proof of principle, for the eventual optimization distribution model, is the development of the expected number of contracts a ZIP produces and the models available for the top five MOSs in the USAR inventory. The derivation of the MOS equations explains approximately 65% of the variation of the data. This is not a perfect model (of course no model is), but it does give explanatory and predictive capability not had previously. Phase I concludes with the determination of the regression equation for the number of contracts a ZIP Code can produce and the top five MOSs in the USAR.

The second thesis, Phase II, in the series will develop models for all 264 MOSs and analyze them for commonalities and differences that reveal insights about recruit production for the USAR. Once we accomplish this for the MOS inventory, we can apply this to the constraint set in Phase III. This will also identify the regional propensity, by using an indicator variable in our regression model, of the market to join the USAR. The third thesis will use those models as constraints in a mixed integer linear program that positions the RCs to maximize their ability to man their units. The assignment of RC market ZIP Codes to maximize unit fill rates leads to increased unit readiness. This thesis creates an initial version of this program.

This thesis automates the process of assembling and reconciling key data files using a commercial data-mining package called Clementine. That process is documented so that future analysts can avoid the nearly three man-months of work it took to create the master data file with its over 30,000 by 430 cells. This is a major contribution.

These results support the solution of the unit fill rate problem and address many of the issues associated with determining the appropriate demographic, economic, and vocational factors of RC markets. Together these three theses will provide a powerful

tool for analysis of optimal reserve force stationing. This will greatly improve the readiness of the Reserve Components, unit deployment schedules, and Homeland Security.

B. CONCLUSIONS

This thesis assembled a database of recruiting, demographic, and economic data by ZIP Code. This database enables the modeling of potential recruit production by ZIP Code for the USAR. Since members of the USAR must in general live within 75 miles or 90 minutes of their RC, ZIP Code level detail is important for understanding the capability of a region to support its reserve units.

The assembly of this data set was a difficult task. The thesis outlines the challenges, and more importantly, preserves the data mining algorithms developed in Clementine so that the next analyst's work can be greatly reduced.

The thesis developed regression models to predict the expected number of contracts that a ZIP Code could produce, and upper bounds for the number of those contracts that could be assigned to five representative MOSs. These expected values and bounds by ZIP Code can be developed for all 264 MOSs and 30,000 ZIP Codes in the United States, and that is proposed for a subsequent thesis. In turn, those values become constraints for the positioning of reserve units. We develop an LP to address that problem, and it is proposed as a third thesis.

The regression models explain about two-thirds of the variation in recruit production and MOS potential. Remaining variation in recruit production is likely affected by policy variables (such as incentives) not captured in the database. Remaining variation in MOS potential likely reflects the underlying variability of educational attainment in the population.

Some of the lessons learned in the Phase I process are variability of the ZIP Codes in demographic composition, among regions of the country, and across vocational information as well. We set out to find an explanation of the relationship of our data elements. We assumed that vocational, market lifestyle segmentation, unemployment

rates, etc. would be the explanatory variables between recruiting and unit fill. The amount of explained variation in the data is about 70% for contracts and about 65% for the five MOSs constructed.

We demonstrated that ZIP Codes have different quality composition even when the numbers of recruits are similar. This quality aspect of the ZIP Code and subsequent market is the key to getting the right type of unit in the right location. This supportability is paramount to unit fill rates. The outcome of this analysis highlights the importance of considering quality in stationing decisions.

As previously stated, there may be something not captured in the data. This may be the periodicity of the data. This phenomenon could be explored to ascertain whether times series is an appropriate model of consideration. There may be seasonality, trend, and other information that was not captured in our developed model.

Subsequent analysis may be able to capture additional information in a time series and subsequently use these forecasts to incorporate them into a better predictive model. The time series alternative should be explored to ascertain whether it might prove to be more beneficial. Now that the data streams are complete, the analytical data runs are an automated process making it easier to update the data. All it takes now is time to complete the stream runs in Clementine. The effort for Phase II can be concentrated on the model for each of the MOSs.

These results support the unit fill rate problem and address many of the issues associated with determining the appropriate demographic, economic, and vocational factors of RC markets. When combined with Phase II and Phase III the model in its entirety will greatly contribute to unit personnel and training readiness. This will greatly aid in the reliance on the Reserve Components, unit deployment schedules, and Homeland Security.

We can and will provide the strength, fill the ranks, train and lead our units to be the best combat multiplier in the world, today, tomorrow, and in the future.

C. RECOMMENDATIONS

I principally recommend that OCAR pursue the completion of the two successor theses outlined in this thesis, so that unit fill potential is included in the discussion of positioning of reserve units. This is particularly timely as the nation prepares for another round of BRACs in 2005.

I also recommend that OCAR construct a data-warehouse that automates the collection of the data using the methods in this thesis, and that automatically reconciles the discrepancies discovered in this thesis. It would be an easy task to assign to a contractor, and would greatly improve the ability of the entire USAREC analyst community to model local effects on recruit production.

These models explain about 70% of the variation in recruit production. This demonstrates the effectiveness of regression and its predictive nature. Phase II needs to continue to pursue the number of contracts and the MOS build. I recommend exploration of the use of times series to explore the MOS models. Phase I results have predictive power, but there may be other factors that will explain additional variation in the data.

Currently each RC has associated market ZIP Codes. I recommend this process to determine those ZIP Codes more appropriate for the current force structure or give insight as to the type of force structure best supported by the market. In each case, we can derive through the analysis the appropriate MOS, vocational, or lifestyle segmentation aspects for each RC.

It would also be advisable to ensure that future studies, structure placement initiatives, recruiter placement initiatives, and any other initiatives be succinctly coordinated between the USARC, USAREC, and a Joint Partnership Evaluation Team responsible for ensuring transitional initiatives are planned, coordinated, and executed in unison for America's Army.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A: TABLE DEFINITIONS DICTIONARY

This is the data definition dictionary for the tables used in the analysis of USAR unit fill. This thesis used the following data and tables to determine a correlation between elements and use it to predict the outcome of stationing actions:

TABLE/SOURCE	DEFINITION
FRC_FILE.DBF (74,176 Records) / OCAR	The table has the structure of every unit in the USAR and its authorized, required, assigned strength totals. It will be used to determine the units needing or having fill problems. The UNIT FILL RATE = ASSIGNED STRENGTH / AUTHORIZED (by MOS) for each unit in the USAR.
PM03.DBF (4,778,080 Records) / USAREC	The table has the Military Available population by race, ethnicity, gender, and ZIP code level of detail. The data is for FY 2000-2020 projected with the anticipated growth rates of the population due to trend analysis. Data is current as of FY2003.
MV50.DBF (43,362 Records) / USAREC	The table has the Microvision Lifestyle Segmentation for each ZIP Code. It will determine the most prominent segments in the ZIP code and used to determine correlations among enlistments and MOS Skill sets at the ZIP Code Level, FIP Code, or Reserve Center Level. **(See Appendix D)**
ALLARMY.DBF (459,761 Records) / USAREC	The table has the quality of enlistments for the Army for FY1999-FY2003 by ZIP Code for each applicant who made entry into the USAR. It contains contract data for all components of the Army.
SISSERV.DBF (646,816 Records) / USAREC	The table has Sister Service contract data for FY 1999-2003. It will be used to determine Sister Service competition on a market.
QUALS.DBF (458 Records) / USAREC	The table has the required ASVAB Test Scores, by category, for each MOS in the inventory. Its use will be to determine the minimum required test score for each applicant to obtain an MOS. If the market cannot test sufficiently high enough to obtain an MOS, we conclude the RC may not support the MOS.
P050.DBF (33,178 Records) / USBC	The table has the Bureau of Labor and Statistics Vocational data for each ZIP Code. It contains information by vocations of the working population aged 16-69 in each ZIP Code. This information will be used to determine the most prominent vocation of each ZIP Code to determine a correlation of MOS Skills with the market/ZIP Code.
RCMKT75.DBF (387,872 Records) / USAREC	The table has the market ZIP Codes for each RC. Each market ZIP Code is not unique; it may be a market ZIP for multiple RCs. The market ZIPs are those within 75 miles of each RC.
LAUCNTY.DBF (3,218 Records) / BLS	The table has the Employment and Unemployment Data for each County in the US verified for 2003. This table has the Labor Force, Employed Labor Forced, Unemployed Labor Force, and the Unemployment Rate for each US County. Unemployment Rate = Unemployed / Labor Force
gp.data.1.AllData.DBF (130,904 Records) / BLS	The table has the Employment Data for each State from 1981– 1999. The table has both seasonal and unseasonal data.

gp.state.DBF **(52 Records) / BLS**	The table has the numerical State codes for each of the fifty states plus those for DC and Puerto Rico.

APPENDIX B: TABLE DATA FIELDS AND DESCRIPTIONS

This is the data field and field descriptions for each table used in the analysis of USAR unit fill:

TABLE	FIELD_NAMES	FIELD DESCRIPTION
JOBMV50.DBF **(Derived Table)**	ZIP P01_TOT TOT_yyyyyy M02_MALE Mxx_yyyyyy F49_FEMALE Fxx_yyyyyy TTL_MV50 PCT_Mvxx	➔ ZIP Code for the Data Elements ➔ Total Working Population in ZIP Code ➔ Total Categorical Working Population in ZIP Code (Same as MALE + FEMALE for Category) [yyyyyy] ➔ MGTPRO, BUSFIN, MGTOTH, FRMMGR, BUSFI2, BUSOPS, FINSPC, PRFSNL, CMPMTH, ARCENG. ARCSUR, DRENMA, LPSSCI, CMSOSV, LGLOCC, EDTRLI, ARETSP, HLTPRA, HDITRT, HLTTCH, SVCOCC, HLTSPT, PRTSVC, FFPRLW, PRTOTH, FDPRSV, BLGRCL, PSLSVC, SALOFF, SALOCC, ADMSPT, FMFIFO, CNEXMT, CONEXT, SUPCON, CONTRD, EXTRTN, INMTRP, PRTRMA, PRDOCC, TRMAMV, SUPTRA, ACRATC, VEHOPR, RLWTOT, MTLMOV ➔ Total Male Working Population in ZIP Code ➔ Total Categorical Male Working Population in ZIP Code [xx] ➔ 03-48 [yyyyyy] ➔ Same as previous ➔ Total Female Working Population in ZIP Code ➔ Total Categorical Female Working Population in ZIP Code [xx] ➔ 49-95 [yyyyyy] ➔ Same as previous ➔ Total Count of MV Segments in the ZIP Code ➔ Percentage of MVxx Segment in the ZIP Code
FRC_FILE.DBF	UIC ACTCO EDATE UNMBR STNNMR LOCCO STOFF STWOF STENL AUOFF AUWOF AUENL TIER STACO LASTUPDT FY	➔ The Unit Identification Code ➔ The Activation Code of the pending action ➔ The Effective Date of the pending action ➔ The Unit Number ➔ The Station Number or ZIP Code ➔ The Location Code or State of the unit ➔ The Stationed count of Officers in the unit ➔ The Stationed count of Warrant Officers in the unit ➔ The Stationed count of Enlisted in the unit ➔ The Authorized count of Officers in the unit ➔ The Authorized count of Warrant Officers in the unit ➔ The Authorized count of Enlisted in the unit ➔ The Tier level of the unit ➔ The Station Code of the unit ➔ The Last Date of the entry of information for the unit ➔ The Fiscal Year of the pending action
PM03.DBF	ZIPCODE RACE SEX Y2000 AGE	➔ The ZIP Code of the population information ➔ The Race of the population ➔ The Sex of the population ➔ The Year of the population information (Year range [2000, 2020]) ➔ The Age of the population

TABLE	FIELD_NAMES	FIELD DESCRIPTION
MV50.DBF (See Appendix D)	ZIP Mvxx TTL_MV50 PCT_MVxx	➔ The ZIP Code for the Data Elements ➔ Count of MV Segments in the ZIP Code [xx] ➔ 01-50 ➔ Total Count of MV Segments in the ZIP Code ➔ Percentage of MVxx Segment in the ZIP Code
ALLARMY.DBF	FY SSN AFQT GT GM EL CL MM SC CO FA OF ST RCZIP COMP_CD ZIP SEGMENT UIC MOS RDOE SKILL_LEVEL	➔ The Fiscal Year of the accession action ➔ Individual's Social Security Number ➔ The Armed Forces Qualification Test Score (0-99) ➔ General Technical Categorical ASVAB Line Score ➔ General Mechanical Categorical ASVAB Line Score ➔ Electrical Categorical ASVAB Line Score ➔ Clerical Aptitude Categorical ASVAB Line Score ➔ Mechanical Maintenance Categorical ASVAB Line Score ➔ Signal & Communications Categorical ASVAB Line Score ➔ Combat Operations Categorical ASVAB Line Score ➔ Field Artillery Categorical ASVAB Line Score ➔ Operations & Food Service Categorical ASVAB Line Score ➔ Science & Technology Categorical ASVAB Line Score ➔ The Reserve Center ZIP Code, if any, of the accession action ➔ Component Code (G-Guard, V-Reserve, R-Regular Army) ➔ ZIP Code for Data Element ➔ Microvision Lifestyle Segment (1-50) ➔ Unit Identification Code of the Unit the Individual joined ➔ Military Occupational Specialty Code ➔ The Reserve Date of Enlistment of the accession action ➔ Skill Level of the MOS (1-5)
SISSERV.DBF	M_ZIP SERV_COMP SEX MEP_RACE MEP_ETHIN DOB EDYRS EDLEVEL HEIGHT WEIGHT PUHLES AFQT zz_SCORE TSC MS_FY	➔ ZIP Code for the Data Elements ➔ The Service Component for the accession ➔ The Sex of the accession ➔ The Race of the accession ➔ The Ethnic Code of the accession ➔ The Date of Birth of the accession ➔ The Years of Education of the accession ➔ The Education Level of the accession [0-24] ➔ The Height of the accession ➔ The Weight of the accession ➔ The PUHLES scores from accessioned physical ➔ The AFQT Score of the accession ➔ The Categorical Raw ASVAB Scores for the accession [zz] ➔ GS, AR, WK, PC, NO, CS, AS, MK, MC, EI, and VE ➔ The Test Score Category of the accession ➔ The Fiscal Year of the accession
QUALS.DBF	MOS4 CMF CMF_DESCR VOCATN AFQT GT GM EL CL MM SC CO FA	➔ The 4 Character Military Occupational Specialty (MOS) ➔ Career Management Field (CMF) of the MOS ➔ Description of the Numerical CMF ➔ The BLS Vocation (13 Major Categories) ➔ The Armed Forces Qualification Test Score (0-99) ➔ General Technical Categorical ASVAB Line Score ➔ General Mechanical Categorical ASVAB Line Score ➔ Electrical Categorical ASVAB Line Score ➔ Clerical Categorical ASVAB Line Score ➔ Mechanical Maintenance Categorical ASVAB Line Score ➔ Signal & Communications Categorical ASVAB Line Score ➔ Combat Operations Categorical ASVAB Line Score ➔ Field Artillery Categorical ASVAB Line Score

TABLE	FIELD_NAMES	FIELD DESCRIPTION
	OF ST	➔ Operations & Food Service Categorical ASVAB Line Score ➔ Science & Technology Categorical ASVAB Line Score
P050.DBF	ZIP P01_TOT TOT_yyyyyy M02_MALE Mxx_yyyyyy F49_FEMALE Fxx_yyyyyy	➔ ZIP Code for the Data Elements ➔ Total Working Population in ZIP Code ➔ Total Categorical Working Population in ZIP Code (Same as MALE + FEMALE for Category) [yyyyyy] ➔ MGTPRO, BUSFIN, MGTOTH, FRMMGR, BUSFI2, BUSOPS, FINSPC, PRFSNL, CMPMTH, ARCENG. ARCSUR, DRENMA, LPSSCI, CMSOSV, LGLOCC, EDTRLI, ARETSP, HLTPRA, HDITRT, HLTTCH, SVCOCC, HLTSPT, PRTSVC, FFPRLW, PRTOTH, FDPRSV, BLGRCL, PSLSVC, SALOFF, SALOCC, ADMSPT, FMFIFO, CNEXMT, CONEXT, SUPCON, CONTRD, EXTRTN, INMTRP, PRTRMA, PRDOCC, TRMAMV, SUPTRA, ACRATC, VEHOPR, RLWTOT, MTLMOV ➔ Total Male Working Population in ZIP Code ➔ Total Categorical Male Working Population in ZIP Code [xx] ➔ 03-48 [yyyyyy] ➔ Same as previous ➔ Total Female Working Population in ZIP Code ➔ Total Categorical Female Working Population in ZIP Code [xx] ➔ 49-95 [yyyyyy] ➔ Same as previous
RCMKT75.DBF	RCZIP MKTZIP	➔ The RC ZIP Code ➔ A Market ZIP Code of the RC. ZIP Codes are within a 75-mile radius of the RC.
LAUCNTY.DBF	LAUS_CODE ST_FIPS CNTY_NAME ST_NAME ST_ABBR YEAR LBR_FRC EMPL UNEMPL UNEMPL_RATE	➔ The Local Area Unemployment Code ➔ The State FIPS used by BLS and USBC. (Same as gp.state.DBF) 01=Alabama, 02=Alaska, …, 56=Wyoming ➔ The County Name ➔ State Name for each State used in the table. ➔ 2 Letter State Abbreviation as provided by BLS and USBC. ➔ The Year of the information. ➔ Labor Force Population in the County. ➔ Employed Labor Force in the County. ➔ Unemployed Labor Force in the County. ➔ Unemployment Rate for the County. (UNEMPL/LBR_FRC)
gp.data.1.AllData .DBF	SERIES_ID YEAR PERIOD VALUE FOOTNOTE	➔ The Series Identification Number (GPU00100000E0000) ➔ The Year of the Data ➔ The Period of the Data ➔ The Value of the Data ➔ The Footnote Codes of the Data (Variable Information) The series_id (GPU00100000E0000) can be broken out into: survey abbreviation=GP, seasonal (code) =U, area_type_code =0, state_code =01, area_code=0000, labor_force_code=E, charact_code=0000
gp.state.DBF	STATE_CODE	➔ State Code used by BLS and USBC. 01=Alabama, 02=Alaska, 04=Arizona, 05=Arkansas, 06=California, 08=Colorado, 09=Connecticut, 10=Delaware, 11=D.C., 12=Florida, 13=Georgia, 15=Hawaii, 16=Idaho, 17=Illinois, 18=Indiana, 19=Iowa, 20=Kansas, 21=Kentucky, 22=Louisiana, 23=Maine, 24=Maryland, 25=Massachusetts,

TABLE	**FIELD_NAMES**	**FIELD DESCRIPTION**
		26=Michigan, 27=Minnesota, 28=Mississippi, 29=Missouri, 30=Montana, 31=Nebraska, 32=Nevada, 33=New Hampshire, 34=New Jersey, 35=New Mexico, 36=New York, 37=North Carolina, 38=North Dakota, 39=Ohio, 40=Oklahoma, 41=Oregon, 42=Pennsylvania, 44=Rhode Island, 45=South Carolina, 46=South Dakota. 47=Tennessee, 48=Texas, 49=Utah, 50=Vermont, 51=Virginia, 53=Washington, 54=West Virginia, 55=Wisconsin, 56=Wyoming

APPENDIX C: OCCUPATIONS AND WORKING CLASS CATEGORIES

<u>White Collar Category Occupations</u>

Executive and Managerial: [EXECMNGE]
Legislators
Chief Executives and General Administrators, Public Administration
Administrators and Officials, Public Administration
Administrators, Protective Services
Financial Managers
Personnel and Labor Relations Managers
Purchasing Managers
Managers, Marketing, Advertising, and Public Relations
Administrators, Education and Related Fields
Managers, Medicine and Health
Managers, Properties and Real Estate
Postmasters and Mail Superintendents
Funeral Directors
Managers and Administrators
Management Related Occupations

Professional Specialty: [PROFSNL]
Mathematical and Computer Scientists
Natural Scientists
Architecture and Engineering Occupations
Architects, Surveyors, Cartographers, and Engineers
Health Diagnosing Occupations
Health Assessment & Treating Occupations
Teachers, Post-secondary
Teachers, except Post-secondary
Counselors, Educational and Vocational Librarians, Archivists, and Curators
Social Scientists and Urban Planners
Social, Recreation, and Religious Workers

Technical Support: [TECHSPT]
Health Technologists and Technicians
Technologists & Technicians, except Health
Drafters, Engineering, and Mapping Technicians
Science Technicians
Technicians, except Health, Engineering, and Science

Sales Occupations: [SALES]
Supervisors and Proprietors
Sales Occupations
Sales Representatives
Commodities except Retail
Sales Workers, Retail and Personal Services and Sales Related Occupations

Administrative Support: [ADMINSPT]
Supervisors
Administrative Support Occupations
Computer Equipment Operators
Secretaries, Stenographers, and Typists
Information Clerks
Records Processing Occupations, except Financial
Financial Records Processing Occupations
Duplicating, Mail & Other Office Machine Operators
Communications Equipment Operators
Mail and Message Distributing Occupations
Material Recording, Scheduling, and Distributing Clerks
N.E.C.
Adjusters and Investigators
Miscellaneous Administrative Support Occupations

Blue Collar Category Occupations

Farm, Forestry & Fish: [FAFOFISH]
Farm Operators and Managers
Other Agricultural and Related Occupations
Forestry and Logging Occupations
Fishers, Hunters, and Trappers

Laborers: [LABORERS]
Supervisors, Handlers, Equipment Cleaners Helpers, Mechanics and Repairers
Helpers, Construction and Extractive Occupations Construction Laborers
Production Helpers
Freight Stock and Materials Handlers
Garage and Service Station, Related Occupations
Vehicle Washers and Equipment Cleaners
Hand Packers

Other Service (except Protective & Household): [SVCOTHR]
Arts, Design, Entertainment, Sports, and Media Occupations
Food Service Preparation and Service Occupations
Health Service Occupations
Cleaning and Building Service Occupations, except Household
Personnel Service Occupation
Launderers and Ironers
Cooks, Private Household
Housekeepers and Butlers
Childcare Workers, Private Households Private Household Cleaners and Servants

Precision Craftsmen: [CRFTSMAN]
Mechanics and Repairers
Construction Trades
Construction Trades, except Supervisors
Extractive Occupations
Precision Production Occupation
Precision Woodworking
Precision Textile, Apparel, and Furnishings Machine Operators
Precision Food Production
Precision Inspectors, Testers, and Related Workers
Plant and System Operators
Metal Working and Plastic Working Machine Operators Fabricating Machine Operators
Metal and Plastic Processing Machine Operators Woodworking Machine Operators
Printing Machine Operators
Textile, Apparel, and Furnishing Operators Machine Operators, Assorted Materials

Protective Service: [SVCPROT]
Supervisors, Protective Service Occupation
Firefighting and Fire Prevention
Police and Detectives
Guards

Transportation & Material Moving: [TRANSPO]
Aircraft and Traffic Control Operators
Motor Vehicle Operators
Transportation Occupations, except Motor Vehicles
Railroad Transportation
Water Transportation
Material Moving Equipment Operators
Production, Transportation, and Material Moving Occupations
Operating Engineers
Long Shore
Hoist & Winch Operators Crane & Tower Operators

P050 TABLE NUMBER & DESCRIPTION

MALE	FEMALE	DESCRIPTION	CATEGORY
P050002	P050049	Total in Population	
P050003	P050050	Management, professional, and related occupations	EXECMNGE
P050004	P050051	Management, business, and financial operations occupations	EXECMNGE
P050005	P050052	Management occupations, except farmers and farm managers	EXECMNGE
P050006	P050053	Farmers and farm managers	FAFOFISH
P050007	P050054	Business and financial operations occupations	EXECMNGE
P050008	P050055	Business operations specialists	ADMINSPT
P050009	P050056	Financial specialists	ADMINSPT
P050010	P050057	Professional and related occupations	PROFSNL
P050011	P050058	Computer and mathematical occupations	PROFSNL
P050012	P050059	Architecture and engineering occupations	PROFSNL
P050013	P050060	Architects, surveyors, cartographers, and engineers	PROFSNL
P050014	P050061	Drafters, engineering, and mapping technicians	TECHSPT
P050015	P050062	Life, physical, and social science occupations	PROFSNL
P050016	P050063	Community and social services occupations	PROFSNL
P050017	P050064	Legal occupations	PROFSNL
P050018	P050065	Education, training, and library occupations	PROFSNL
P050019	P050066	Arts, design, entertainment, sports, and media occupations	SVCOTHR
P050020	P050067	Healthcare practitioners and technical occupations	TECHSPT
P050021	P050068	Health diagnosing and treating practitioners and technical occupations	PROFSNL
P050022	P050069	Health technologists and technicians	TECHSPT
P050023	P050070	Service occupations	SVCOTHR
P050024	P050071	Healthcare support occupations	TECHSPT
P050025	P050072	Protective service occupations	SVCPROT
P050026	P050073	Fire fighting, prevention, and law enforcement workers, including supervisors	SVCPROT
P050027	P050074	Other protective service workers, including supervisors	SVCPROT
P050028	P050075	Food preparation and serving related occupations	SVCOTHR
P050029	P050076	Building and grounds cleaning and maintenance occupations	SVCOTHR
P050030	P050077	Personal care and service occupations	SVCOTHR

MALE	FEMALE	DESCRIPTION	CATEGORY
P050031	P050078	Sales and office occupations	SALES
P050032	P050079	Sales and related occupations	SALES
P050033	P050080	Office and administrative support occupations	ADMINSPT
P050034	P050081	Farming, fishing, and forestry occupations	FAFOFISH
P050035	P050082	Construction, extraction, and maintenance occupations	CRFTSMAN
P050036	P050083	Construction and extraction occupations	CRFTSMAN
P050037	P050084	Supervisors, construction and extraction workers	LABORERS
P050038	P050085	Construction trades workers	CRFTSMAN
P050039	P050086	Extraction workers	CRFTSMAN
P050040	P050087	Installation, maintenance, and repair occupations	CRFTSMAN
P050041	P050088	Production, transportation, and material moving occupations	TRANSPO
P050042	P050089	Production occupations	TRANSPO
P050043	P050090	Transportation and material moving occupations	TRANSPO
P050044	P050091	Supervisors, transportation and material moving workers	TRANSPO
P050045	P050092	Aircraft and traffic control occupations	TRANSPO
P050046	P050093	Motor vehicle operators	TRANSPO
P050047	P050094	Rail, water and other transportation occupations	TRANSPO
P050048	P050095	Material moving workers	TRANSPO

NOTE: Tables and Descriptions provided by the US Bureau of the Census

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D: MICROVISION 50 LIFESTYLE SEGMENTS

SEG #	SEGMENT NAME	SEGMENT DESCRIPTION	GRP #	GROUP NAME
1	Upper Crust	Metropolitan couples and families, very high income and education, homeowners, very high property values, managers/ professionals	1	Accumulated Wealth
2	Lap of Luxury	Families, teens, very high income and education, homeowners, managers/ professionals, 2-worker families	1	Accumulated Wealth
3	Established Wealth	School-age families, high income. high education, homeowners, managers and professionals	1	Accumulated Wealth
4	Mid-Life Success	Families with high education, high income, managers/professionals, technical/sales	1	Accumulated Wealth
5	Prosperous Metro Mix	Families with young children, high education, high income, managers/professionals, technical/sales	1	Accumulated Wealth
6	Good Family Life	Families, children age 5-17, very high education, high income, executives, managers/professionals, technical/sales, home owners	1	Accumulated Wealth
7	Comfortable Times	Middle-aged heads of household, families, high income, medium-high education, technical/sales, managers/professionals	6	Conservative Classics
8	Movers and Shakers	Singles and couples, students and recent graduates, high education and income, managers/professionals, technical/sales	4	Mainstream Singles
9	Building a Home Life	School-age families, new housing, medium-high education, technical/sales, managers/professionals	3	Young Accumulators
10	Home Sweet Home	Married Couples, one or no children, some retirees, medium-high income and education, managers/ professionals, technical/sales	2	Mainstream Families
11	Family Ties	Large families, medium education, medium-high income, technical/sales, Precision/crafts, two workers	2	Mainstream Families
12	A Good Step Forward	Mobile singles, high education, medium income, often renters, managers/professionals, technical/sales	4	Mainstream Singles
13	Successful Singles	Urban areas, renters, young singles and couples, older housing, ethnic	9	Sustaining Singles

SEG #	SEGMENT NAME	SEGMENT DESCRIPTION	GRP #	GROUP NAME
		mix, high education, medium income, managers/ professionals		
14	Middle Years	Mid-life couples, families, medium-high education, mixed occupations. medium income	1	Accumulated Wealth
15	Great Beginnings	Young, singles and couples, medium-high education, medium income, some renters, managers/professionals, technical/sales	4	Mainstream Singles
16	Country Home Families	Large families, rural areas, medium education, medium income, precision/crafts - trades	2	Mainstream Families
17	Stars and Stripes	Young heads of household, large families with school-age children, medium income and education, some military, precision/craft	2	Mainstream Families
18	White Picket Fence	Young families, low to medium education, medium income, precision/crafts, laborers	2	Mainstream Families
19	Young and Carefree	Young, singles and couples, no kids, medium income, medium-high education technical/sales, managers/ professionals	3	Young Accumulators
20	Secure Adults	Mature/seniors, metro fringe areas, singles and couples, medium income, medium education, mixed occupations and some retirees	6	Conservative Classics
21	American Classics	Seniors, singles and couples, no kids, suburban areas, medium income, medium education, mixed occupations and some retirees	6	Conservative Classics
22	Traditional Times	Seniors, no kids, low education levels, medium income, laborers, precision/crafts workers, some retirees	2	Mainstream Families
23	Settled In	Empty nesters, no kids, medium education and income, some retirees, technical/sales and service occupations	2	Mainstream Families
24	City Ties	School-age families, urban areas, African-American, average income, average education, service and laborer occupations	8	Sustaining Families
25	Bedrock America	School-age families, medium income, low-medium education, precision/crafts, military, laborers	3	Young Accumulators
26	The Mature Years	Couples and small families, medium income, low-medium education, precision/crafts, laborers	7	Cautious Couples
27	Middle of the	School-age families, medium income,	5	Asset-

SEG #	SEGMENT NAME	SEGMENT DESCRIPTION	GRP #	GROUP NAME
	Road	mixed education levels, mixed education levels, mixed occupations		Building Families
28	Building a Family	Families, school-age children, medium income, medium-low education, mixed occupations	3	Young Accumulators
29	Establishing Roots	Families with kids of all ages, medium income, low education. mixed occupations	5	Asset-Building Families
30	Domestic Duos	Mature/seniors, singles and couples, no kids, medium-low income, mixed housing, medium education, technical/sales, managers/professionals, some retirees	6	Conservative Classics
31	Country Classics	Middle-aged to mature heads of household, seniors, medium-low income, low education, some mobile homes, laborers	6	Conservative Classics
32	Metro Singles	Singles, renters, urban areas, multi-unit housing, low education, medium-low income, technical/sales, laborers	4	Mainstream Singles
33	Living Off the Land	Rural areas, school-age families, medium-low income, low education, farming/fishing, laborers	7	Cautious Couples
34	Books and New Recruits	Young, high education, medium-low income, students, managers/professionals, service occupations, some military, renters	4	Mainstream Singles
35	Buy American	Families with school-age kids, medium-low income, low education, laborers	2	Mainstream Families
36	Metro Mix	Young singles, no kids, ethnic mix, medium-low income, mostly renters, multi-unit housing, use public transportation	9	Sustaining Singles
37	Urban Up and Comers	Young, singles, ethnic mix, renters, multi-unit housing, high education, medium-low income, managers/professionals	9	Sustaining Singles
38	Rustic Homesteaders	Rural areas, families, school-age kids, low education, medium-low income, some mobile homes, farming/fishing, laborers	2	Mainstream Families
39	On Their Own	Mix of young and seniors, singles and couples, medium-low income, medium-high education, managers/professionals, technical/sales, some renters	4	Mainstream Singles
40	Trying Metro Times	Mix of young and seniors, urban, ethnic mix, low income, older housing, owners and renter, low education levels, varied occupations.	4	Mainstream Singles

SEG #	SEGMENT NAME	SEGMENT DESCRIPTION	GRP #	GROUP NAME
41	Close Knit Families	Primarily Hispanic, large families, kids of all ages, low income and education, precision/craft occupations and laborers	8	Sustaining Families
42	Trying Rural Times	Large families, ethnic mix, low income and education, some mobile homes, service occupations, laborers	8	Sustaining Families
43	Manufacturing USA	Largely African American, singles and families, older housing, low income and education, service and laborer occupations	8	Sustaining Families
44	Hard Years	Young adults and seniors, low income and education, older multi-unit housing, renters service occupations, laborers	8	Sustaining Families
45	Struggling Metro Mix	Young, singles, urban, cultural mix, renters, low income, mixed education levels, older multi-unit housing	9	Sustaining Singles
46	Difficult Times	Primarily African-American, school-age families, urban areas, very low income, low education, laborers and service occupations	8	Sustaining Families
47	University USA	Students and singles, dorms and group quarters, very low income,-medium-high education, technical/sales	9	Sustaining Singles
48	Urban Singles	Mix of young and seniors, singles, renters, old multi-unit housing, urban areas, very low income, mixed education levels, service occupations, technical/sales	9	Sustaining Singles
49	Anomalies	No homogeneity	10	Anomalies
50	Unclassified	Post Office Boxes and unclassified population	11	Unclassified

APPENDIX E: CLEMENTINE SCREEN SNAPSHOTS

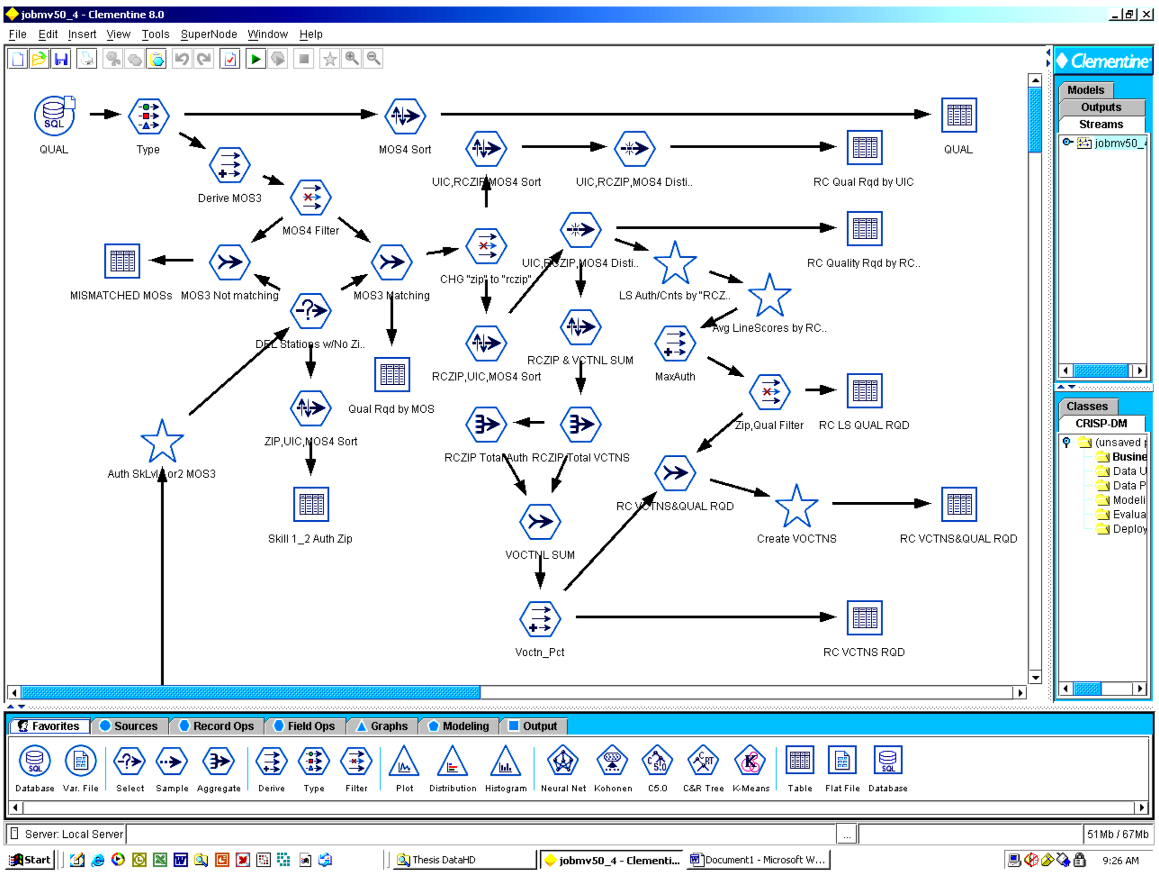


Figure E.1: Clementine Screen Snapshot – QUAL Data Collection

NOTE: All data streams created in Clementine have been saved to a file for future works (Phase II and III). Copies were distributed to my Thesis Advisor: Dr David H. Olwell and Second Reader: Dr Samuel E. Buttrey. These files are also available by request from the author for follow-on analysis.

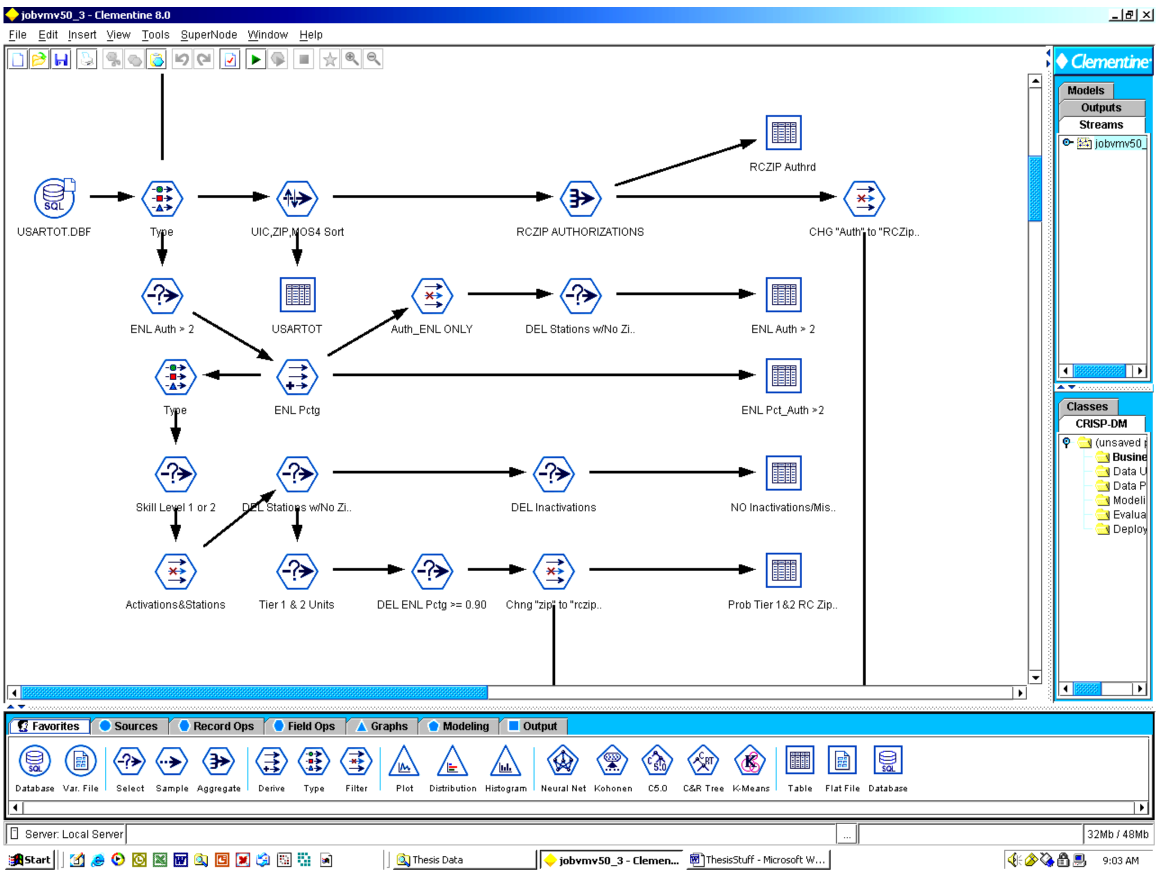


Figure E.2: Clementine Screen Snapshot – USARTOT Data Collection

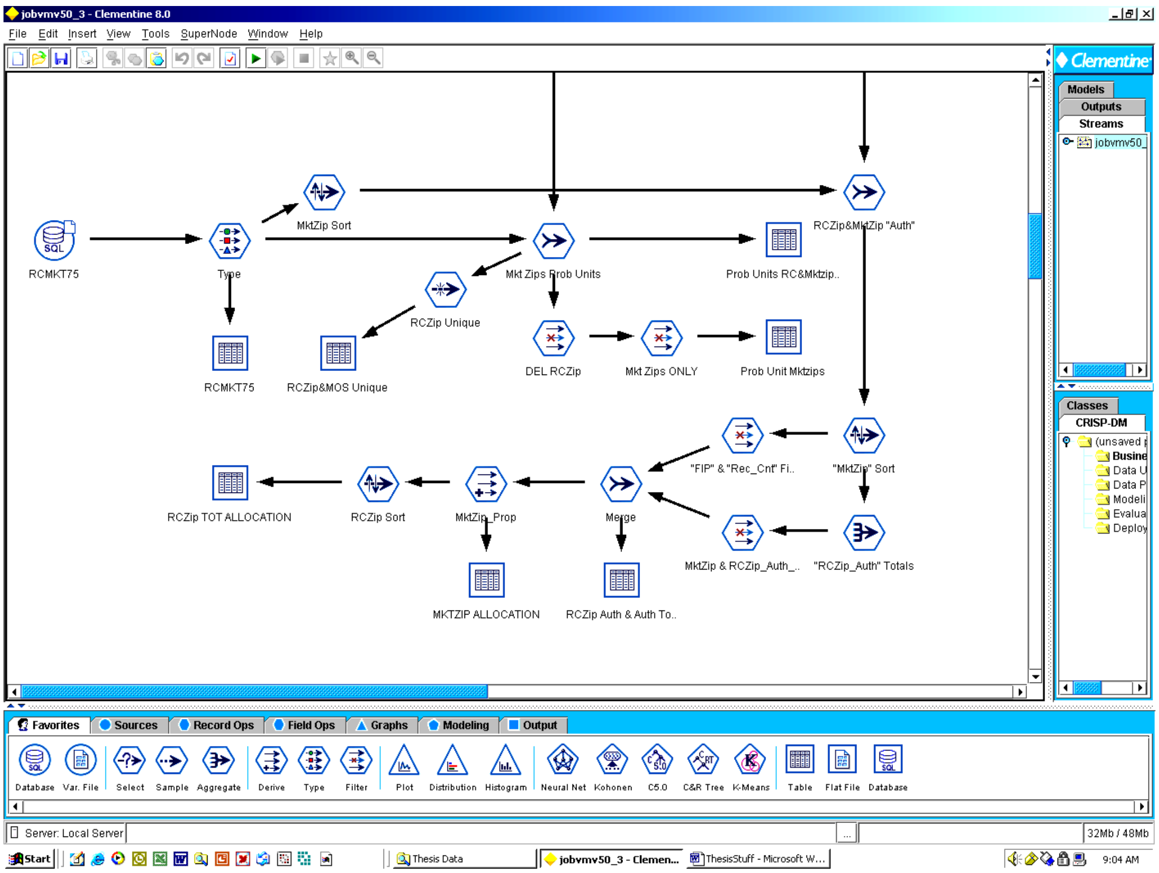


Figure E.3: Clementine Screen Snapshot – RCMKT75 Data Collection

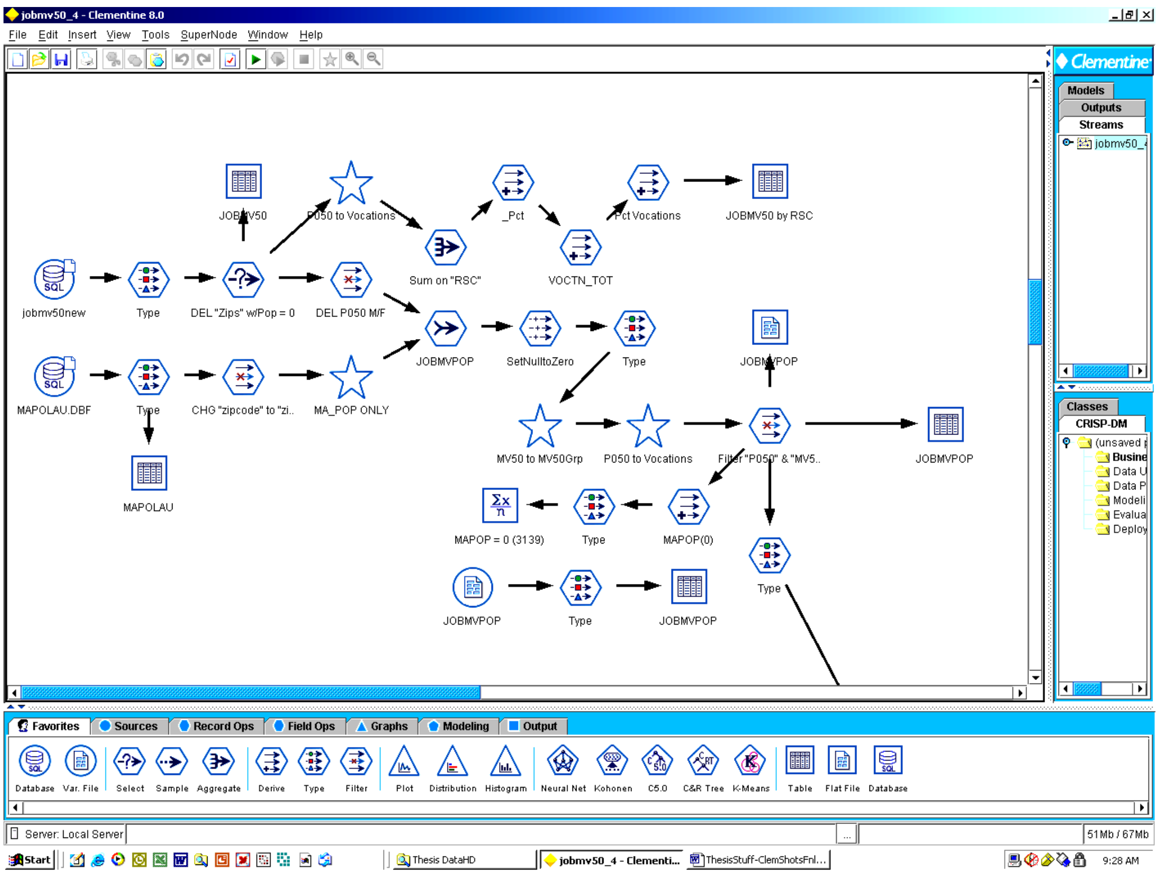


Figure E.4: Clementine Screen Snapshot – JOBMV50 & MAPOPLAU Data Collection

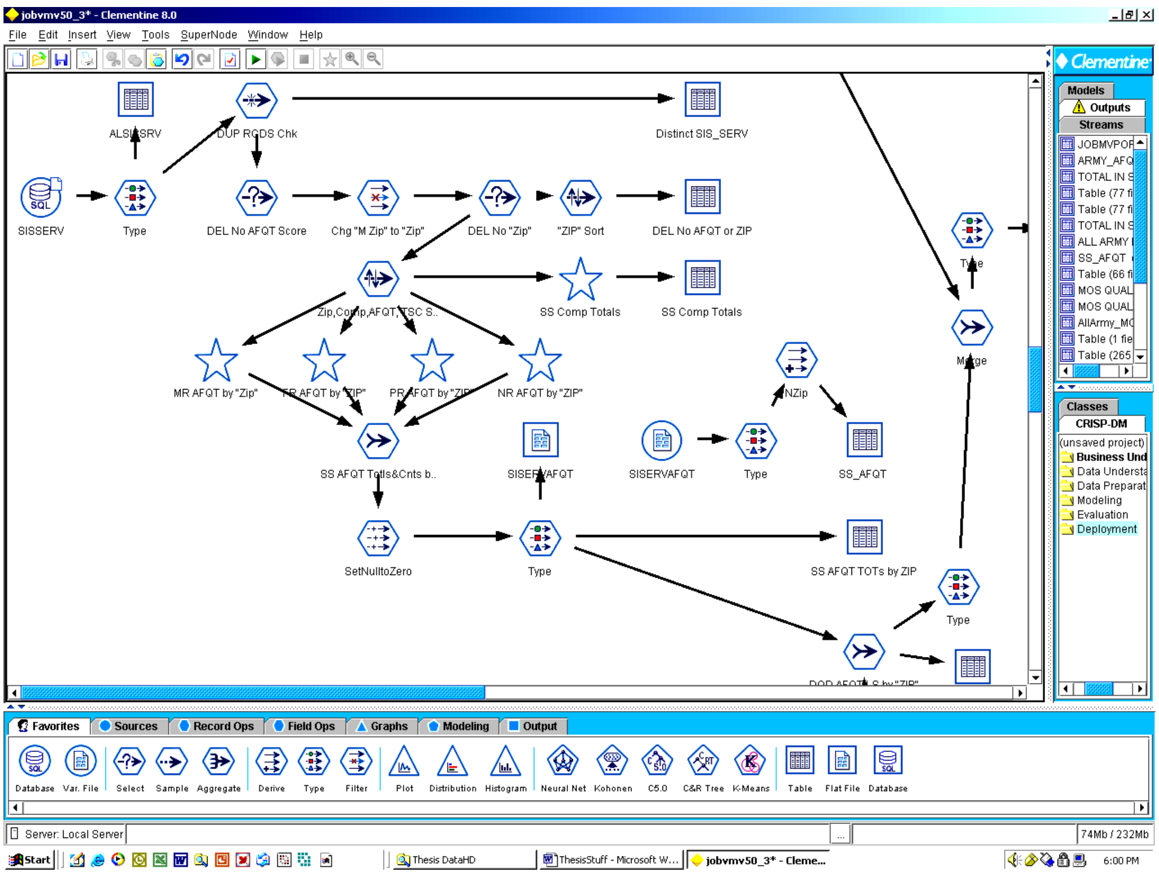


Figure E.5: Clementine Screen Snapshot – SISSERV Data Collection

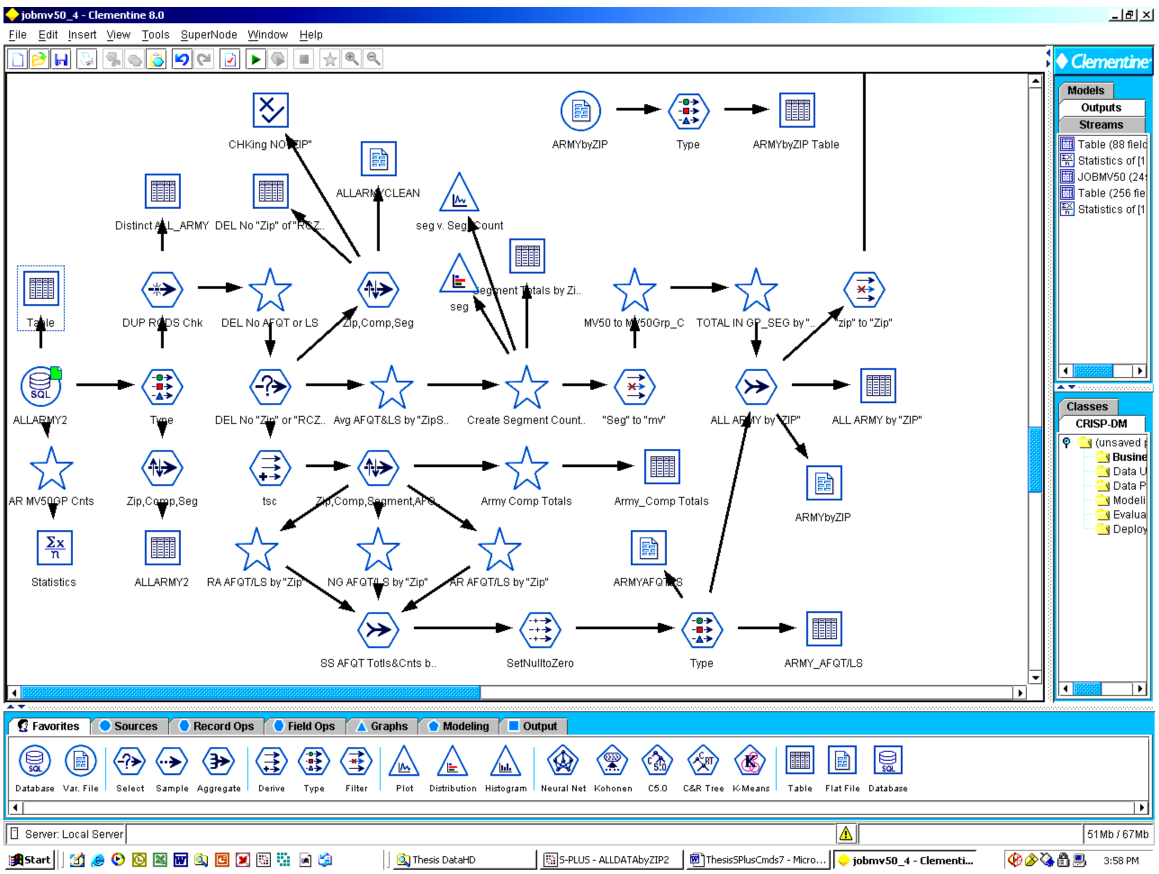


Figure E.6: Clementine Screen Snapshot – ALLARMY (Part 1 of 2) Data Collection

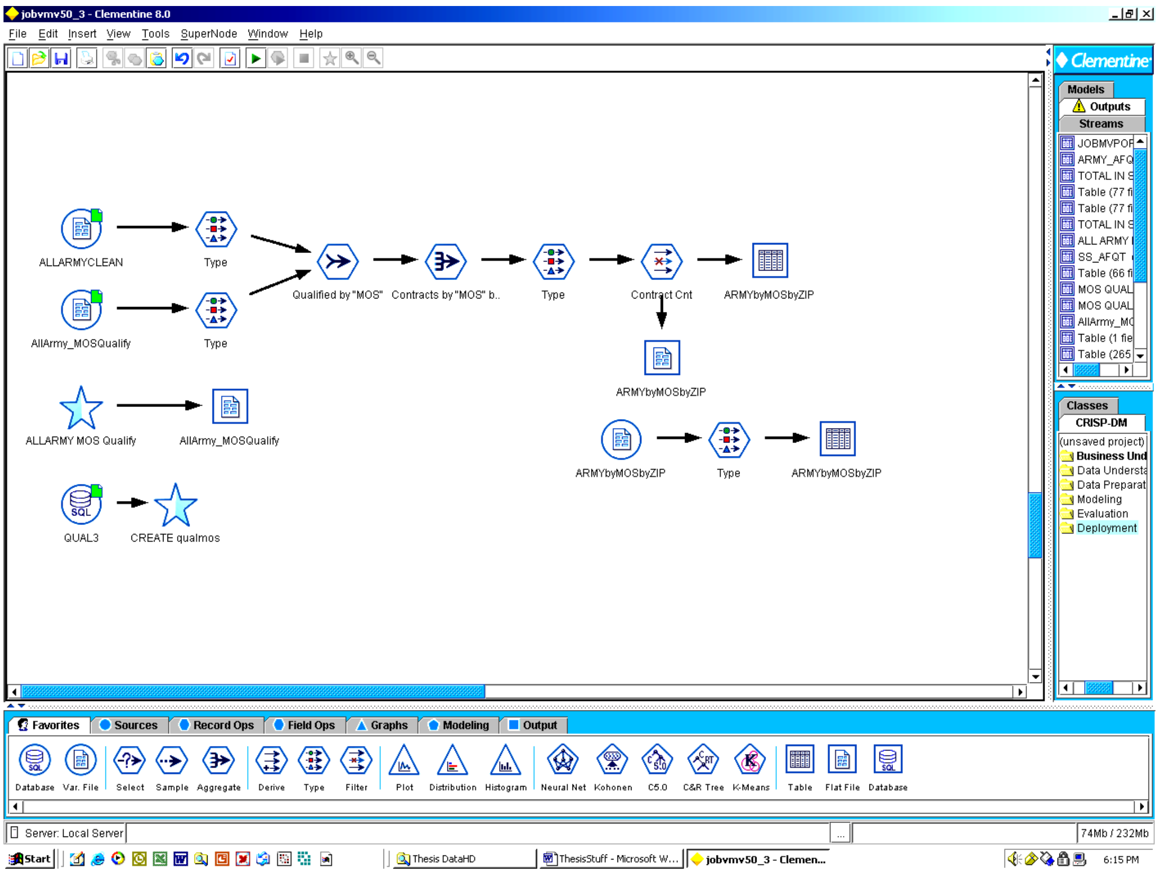


Figure E.7: Clementine Screen Snapshot – ALLARMY(Part 2 of 2) Data Collection

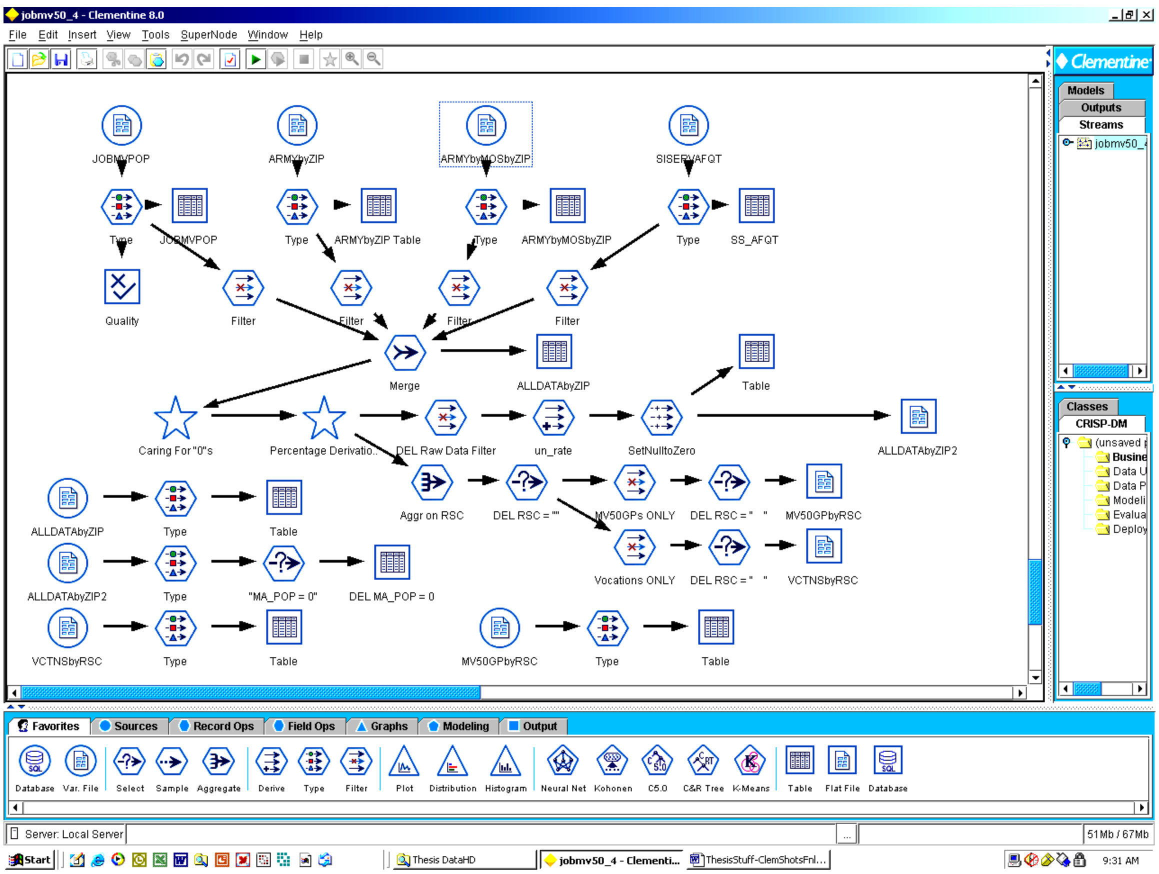


Figure E.8: Clementine Screen Snapshot – ALLDATAbyZIP Data Collection

APPENDIX F: DATA TABLE DERIVATION

Derived tabular information produced by Clementine streams. Appendix E (Clementine Screen Snapshots) contains the graphical representation of the information. Tables derived from collected data contain the following information:

TABLE	DERIVATION
RC VCTNS&QUAL RQD	Produced by merging the USARTOT structure by MOS information and the MOS QUAL table
RCZip TOT ALLOCATION	Produced by merging the USARTOT structure by MOS information and the RCMKT75 table
JOBMVPOP	Produced by merging the JOBMV50new table and the MAPOLAU table. The MAPOLAU table has the BLS Vocational, MA population, and the Local Area Unemployment statistics.
SISERVAFQT	Produced by building the Sister Service component AFQT information
ARMYbyZIP	Produced by building the Army component AFQT information, LSCAT information, MV50 Segmentation information, and MOS Qualification by ZIP Code information. Subsequently merging the three separate pieces of information.
ARMYbyMOSbyZIP	Produced by conducting a quality check of each MOS with contract LSCAT data. Each MOS by ZIP Code was compared to the LSCAT of the contract. If the contract LSCAT $\geq$ MOS needed LSCAT then the contract qualified for the MOS, otherwise it did not.
ALLDATAbyZIP	Produced by merging the JOBMVPOP, SISERVAFQT, ARMYbyZIP, and ARMYbyMOSbyZIP information.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX G: TOP FIVE MOS REGRESSION EQUATIONS

<table>
<tr><th>MOS</th><th>LINEAR REGRESSION MODEL FORMULATION</th></tr>
<tr><td>52D</td><td>

FULL MODEL:

```
q.52D.Avg.Annl ~ un.rate + MA.POP + EXECMNGE + FAFOFISH +
ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES +
CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 +
MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 +
MV50GP10 + MV50GP11


      Residuals:   Min     1Q    Median    3Q   Max
                 -3.89 -0.1487 -0.05103 0.1034 10.6


      Coefficients  Value   Std.Error  t-value   Pr(>|t|)
      (Intercept)   0.1070   0.0093    11.4496   0.0000
          un.rate  -1.2145   0.1402    -8.6628   0.0000
           MA.POP   0.0001   0.0000    24.7223   0.0000
         EXECMNGE  -0.0001   0.0000   -16.5890   0.0000
         FAFOFISH  -0.0003   0.0000   -11.3405   0.0000
         ADMINSPT   0.0002   0.0000    10.5916   0.0000
          PROFSNL   0.0001   0.0000    12.0986   0.0000
          TECHSPT   0.0004   0.0000    17.8766   0.0000
          SVCOTHR   0.0000   0.0000     2.2840   0.0224
          SVCPROT  -0.0001   0.0000    -3.3660   0.0008
            SALES   0.0001   0.0000     8.2252   0.0000
         CRFTSMAN  -0.0001   0.0000   -13.2233   0.0000
         LABORERS  -0.0009   0.0002    -5.2950   0.0000
          TRANSPO   0.0000   0.0000     2.7415   0.0061
         MV50GP01   0.0000   0.0000    -3.1651   0.0016
         MV50GP02   0.0001   0.0000    13.8081   0.0000
         MV50GP03   0.0007   0.0000    33.0811   0.0000
         MV50GP04   0.0000   0.0000   -11.2071   0.0000
         MV50GP05  -0.0009   0.0001    -7.0296   0.0000
         MV50GP06  -0.0001   0.0000   -11.7018   0.0000
         MV50GP07   0.0000   0.0002     0.2591   0.7956
         MV50GP08  -0.0001   0.0000   -13.0907   0.0000
         MV50GP09  -0.0001   0.0000   -16.4919   0.0000
         MV50GP10  -0.0013   0.0003    -4.0472   0.0001
         MV50GP11   0.0000   0.0001     0.0380   0.9697


Residual standard error: 0.5539 on 29839 degrees of freedom
Multiple R-Squared: 0.6559
F-statistic: 2370 on 24 and 29839 degrees of freedom, the p-value is 0
1 observations deleted due to missing values
```

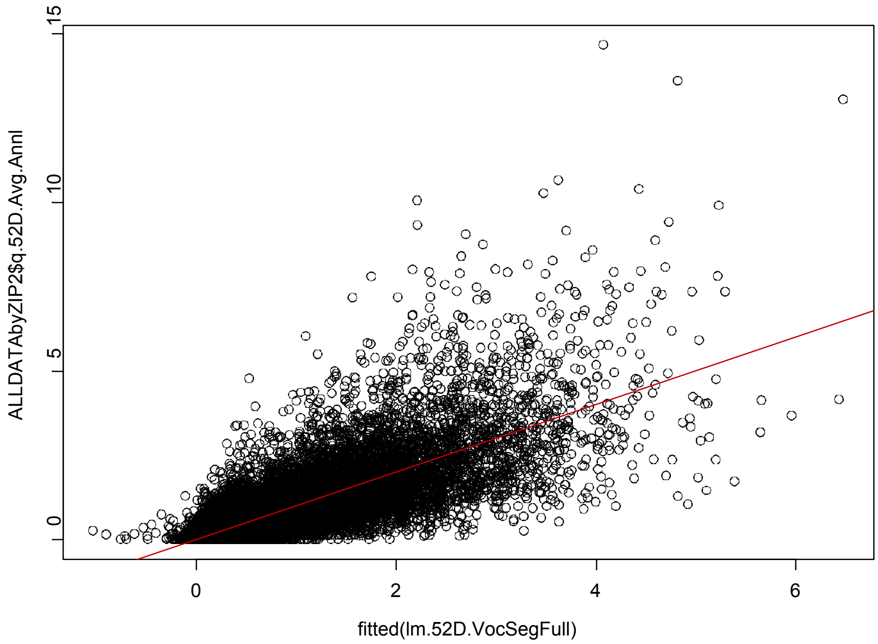
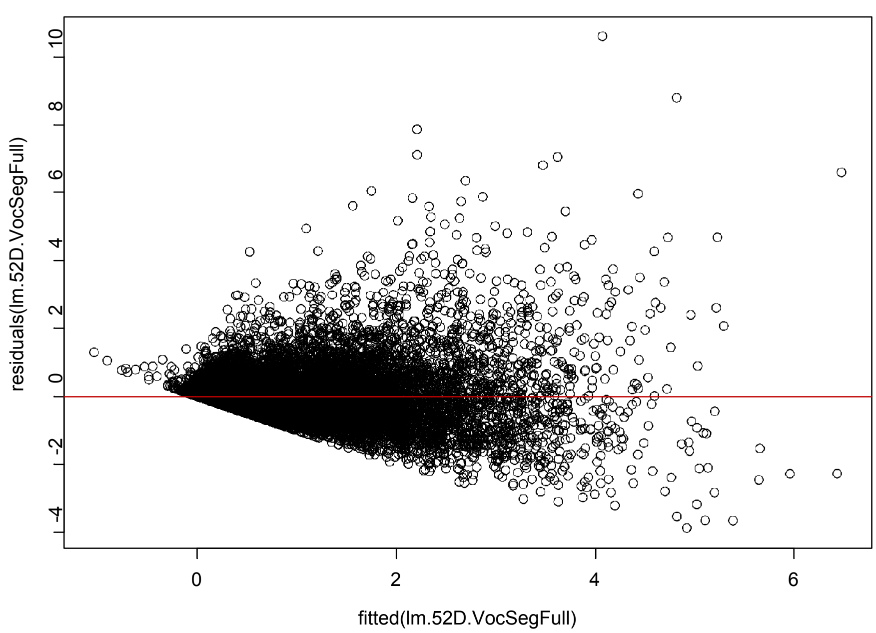
</td></tr>
</table>

MOS	LINEAR REGRESSION MODEL FORMULATION
52D	**FULL MODEL LESS MA.POP and un.rate:** q.52D.Avg.Annl ~ EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-3.82	-0.148	-0.0454	0.1009	10.66

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.0350	0.0046	7.6562	0.0000
EXECMNGE	-0.0002	0.0000	-27.6665	0.0000
FAFOFISH	-0.0002	0.0000	-7.7462	0.0000
ADMINSPT	0.0002	0.0000	10.8476	0.0000
PROFSNL	0.0002	0.0000	26.5587	0.0000
TECHSPT	0.0003	0.0000	12.1059	0.0000
SVCOTHR	0.0001	0.0000	10.0226	0.0000
SVCPROT	-0.0002	0.0000	-5.6266	0.0000
SALES	0.0002	0.0000	14.9044	0.0000
CRFTSMAN	-0.0001	0.0000	-10.7145	0.0000
LABORERS	-0.0010	0.0002	-6.1397	0.0000
TRANSPO	0.0000	0.0000	6.6773	0.0000
MV50GP01	-0.0000	0.0000	-5.9732	0.0000
MV50GP02	0.0000	0.0000	9.8032	0.0000
MV50GP03	0.0008	0.0000	33.5057	0.0000
MV50GP04	-0.0000	0.0000	-10.3981	0.0000
MV50GP05	-0.0011	0.0001	-7.9610	0.0000
MV50GP06	-0.0002	0.0000	-15.0229	0.0000
MV50GP07	0.0003	0.0002	1.5426	0.1229
MV50GP08	-0.0001	0.0000	-10.0891	0.0000
MV50GP09	-0.0001	0.0000	-14.3849	0.0000
MV50GP10	-0.0012	0.0003	-3.5348	0.0004
MV50GP11	0.0000	0.0001	0.0630	0.9498

Residual standard error: 0.5602 on 29842 degrees of freedom
Multiple R-Squared: 0.648
F-statistic: 2497 on 22 and 29842 degrees of freedom, the p-value is 0
1 observations deleted due to missing values

MOS	LINEAR REGRESSION MODEL FORMULATION
52D	 

MOS	LINEAR REGRESSION MODEL FORMULATION
74D	see below

FULL MODEL:

```
q.74D.Avg.Annl ~ un.rate + MA.POP + EXECMNGE + FAFOFISH +
ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES +
CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 +
MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 +
MV50GP10 + MV50GP11
```

Residuals:	Min	1Q	Median	3Q	Max
	-4.522	-0.1587	-0.05132	0.1093	14.16

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.1143	0.0104	11.0120	0.0000
un.rate	-1.3135	0.1558	-8.4324	0.0000
MA.POP	0.0001	0.0000	24.0234	0.0000
EXECMNGE	-0.0001	0.0000	-17.5612	0.0000
FAFOFISH	-0.0004	0.0000	-11.8077	0.0000
ADMINSPT	0.0003	0.0000	14.4229	0.0000
PROFSNL	0.0001	0.0000	10.4161	0.0000
TECHSPT	0.0005	0.0000	19.0585	0.0000
SVCOTHR	0.0000	0.0000	4.1740	0.0000
SVCPROT	0.0000	0.0000	0.5205	0.6027
SALES	0.0001	0.0000	6.4668	0.0000
CRFTSMAN	-0.0001	0.0000	-13.9898	0.0000
LABORERS	-0.0011	0.0002	-6.2129	0.0000
TRANSPO	0.0000	0.0000	1.1808	0.2377
MV50GP01	0.0000	0.0000	-1.4518	0.1466
MV50GP02	0.0001	0.0000	11.2329	0.0000
MV50GP03	0.0009	0.0000	35.1581	0.0000
MV50GP04	0.0000	0.0000	-10.5038	0.0000
MV50GP05	-0.0012	0.0001	-7.8578	0.0000
MV50GP06	-0.0001	0.0000	-12.7840	0.0000
MV50GP07	-0.0003	0.0002	-1.5664	0.1173
MV50GP08	0.0000	0.0000	-6.9234	0.0000
MV50GP09	-0.0001	0.0000	-14.4924	0.0000
MV50GP10	-0.0013	0.0004	-3.6710	0.0002
MV50GP11	0.0000	0.0001	0.1755	0.8607

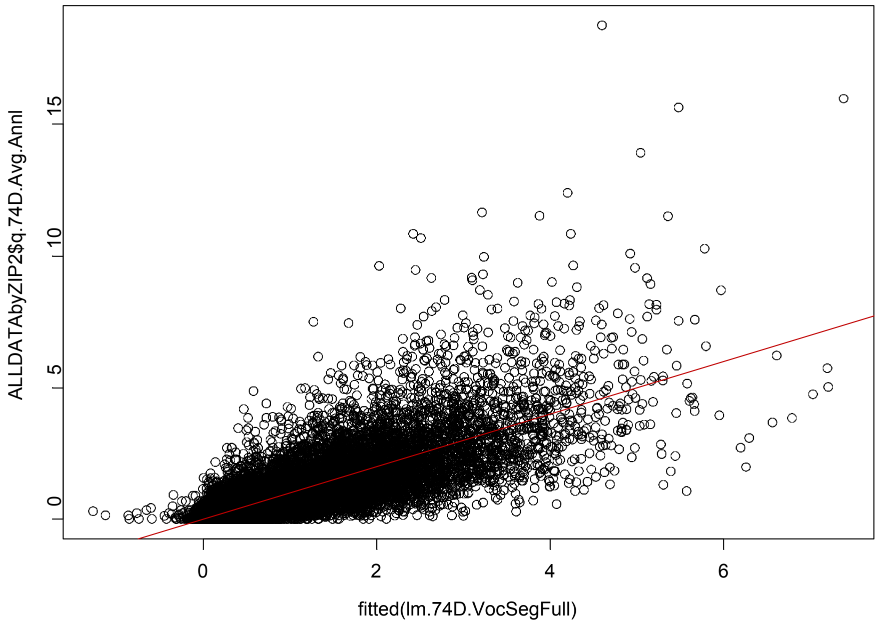
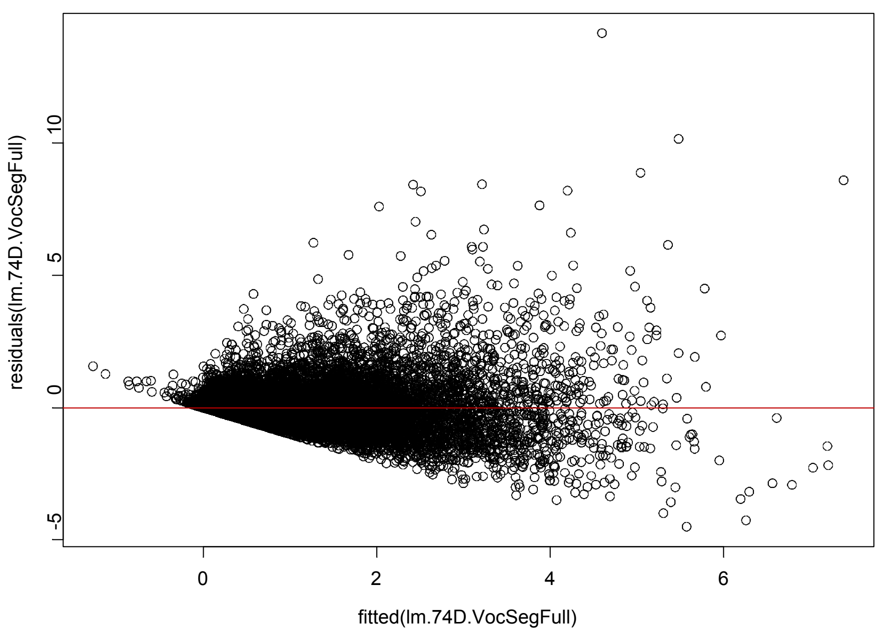
Residual standard error: 0.6154 on 29839 degrees of freedom
Multiple R-Squared: 0.6687
F-statistic: 2509 on 24 and 29839 degrees of freedom, the p-value is 0
1 observations deleted due to missing values

MOS	LINEAR REGRESSION MODEL FORMULATION
74D	**FULL MODEL LESS MA.POP and un.rate:** q.74D.Avg.Annl ~ EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-4.45	-0.1558	-0.0459	0.1068	14.23

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.0365	0.0051	7.1856	0.0000
EXECMNGE	-0.0002	0.0000	-28.4424	0.0000
FAFOFISH	-0.0003	0.0000	-8.3272	0.0000
ADMINSPT	0.0003	0.0000	14.6413	0.0000
PROFSNL	0.0003	0.0000	24.3189	0.0000
TECHSPT	0.0003	0.0000	13.5007	0.0000
SVCOTHR	0.0001	0.0000	11.7635	0.0000
SVCPROT	-0.0001	0.0000	-1.7032	0.0885
SALES	0.0002	0.0000	12.9252	0.0000
CRFTSMAN	-0.0001	0.0000	-11.5517	0.0000
LABORERS	-0.0013	0.0002	-7.0269	0.0000
TRANSPO	0.0000	0.0000	4.9989	0.0000
MV50GP01	-0.0000	0.0000	-4.1921	0.0000
MV50GP02	0.0000	0.0000	7.3438	0.0000
MV50GP03	0.0009	0.0000	35.5627	0.0000
MV50GP04	-0.0001	0.0000	-9.7278	0.0000
MV50GP05	-0.0013	0.0002	-8.7616	0.0000
MV50GP06	-0.0002	0.0000	-16.0188	0.0000
MV50GP07	-0.0001	0.0002	-0.3028	0.7620
MV50GP08	-0.0000	0.0000	-3.9972	0.0001
MV50GP09	-0.0001	0.0000	-12.4566	0.0000
MV50GP10	-0.0012	0.0004	-3.1780	0.0015
MV50GP11	0.0000	0.0001	0.1984	0.8427

Residual standard error: 0.622 on 29842 degrees of freedom
Multiple R-Squared: 0.6615
F-statistic: 2650 on 22 and 29842 degrees of freedom, the p-value is 0

MOS	LINEAR REGRESSION MODEL FORMULATION
74D	 

MOS	LINEAR REGRESSION MODEL FORMULATION
77F	**FULL MODEL:** q.77F.Avg.Annl ~ un.rate + MA.POP + EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-6.16	-0.1916	-0.05805	0.1304	19.26

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.1305	0.0131	9.9687	0.0000
un.rate	-1.4504	0.1964	-7.3848	0.0000
MA.POP	0.0001	0.0000	20.4930	0.0000
EXECMNGE	-0.0001	0.0000	-18.2117	0.0000
FAFOFISH	-0.0005	0.0000	-12.5701	0.0000
ADMINSPT	0.0005	0.0000	18.8855	0.0000
PROFSNL	0.0001	0.0000	7.3996	0.0000
TECHSPT	0.0006	0.0000	20.9363	0.0000
SVCOTHR	0.0001	0.0000	6.2479	0.0000
SVCPROT	0.0002	0.0000	4.3292	0.0000
SALES	0.0001	0.0000	3.5163	0.0004
CRFTSMAN	-0.0002	0.0000	-15.2893	0.0000
LABORERS	-0.0017	0.0002	-7.1630	0.0000
TRANSPO	0.0000	0.0000	1.6578	0.0974
MV50GP01	0.0000	0.0000	0.4156	0.6777
MV50GP02	0.0001	0.0000	8.8137	0.0000
MV50GP03	0.0012	0.0000	40.1701	0.0000
MV50GP04	0.0000	0.0000	-8.6634	0.0000
MV50GP05	-0.0015	0.0002	-8.1599	0.0000
MV50GP06	-0.0002	0.0000	-14.0749	0.0000
MV50GP07	-0.0007	0.0002	-3.1489	0.0016
MV50GP08	0.0000	0.0000	6.1784	0.0000
MV50GP09	-0.0001	0.0000	-11.6317	0.0000
MV50GP10	-0.0014	0.0005	-3.1333	0.0017
MV50GP11	0.0000	0.0001	0.5078	0.6116

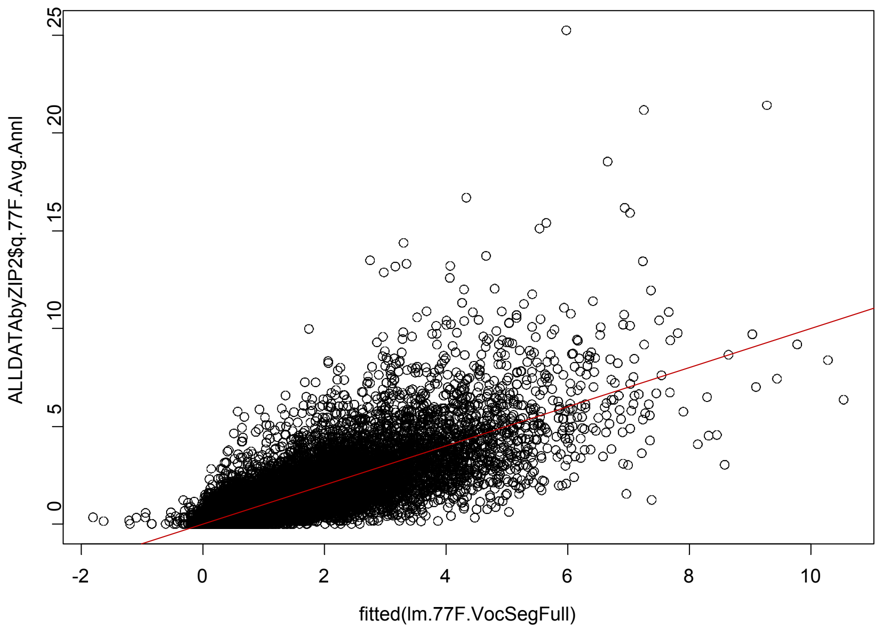
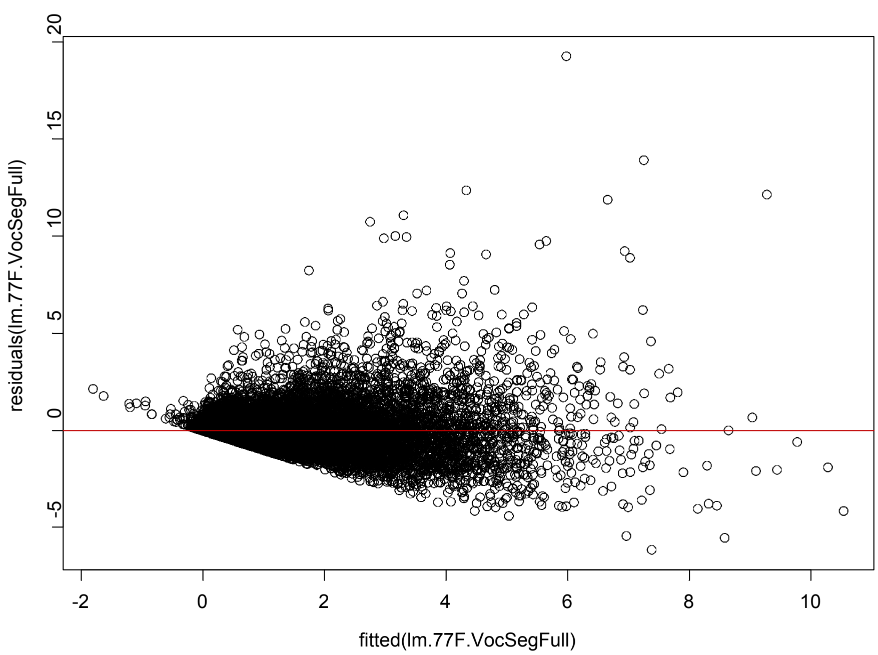
Residual standard error: 0.7759 on 29839 degrees of freedom
Multiple R-Squared: 0.6811
F-statistic: 2656 on 24 and 29839 degrees of freedom, the p-value is 0
1 observations deleted due to missing values

MOS	LINEAR REGRESSION MODEL FORMULATION
77F	**FULL MODEL LESS MA.POP and un.rate:** q.77F.Avg.Annl ~ EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-6.08	-0.1898	-0.0516	0.1244	19.33

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.0446	0.0064	6.9823	0.0000
EXECMNGE	-0.0002	0.0000	-27.7956	0.0000
FAFOFISH	-0.0004	0.0000	-9.6641	0.0000
ADMINSPT	0.0005	0.0000	19.0766	0.0000
PROFSNL	0.0002	0.0000	19.1293	0.0000
TECHSPT	0.0005	0.0000	16.3652	0.0000
SVCOTHR	0.0001	0.0000	12.8605	0.0000
SVCPROT	0.0001	0.0000	2.4031	0.0163
SALES	0.0002	0.0000	8.9922	0.0000
CRFTSMAN	-0.0002	0.0000	-13.2271	0.0000
LABORERS	-0.0018	0.0002	-7.8593	0.0000
TRANSPO	0.0000	0.0000	4.9229	0.0000
MV50GP01	-0.0000	0.0000	-1.9400	0.0524
MV50GP02	0.0000	0.0000	5.5008	0.0000
MV50GP03	0.0013	0.0000	40.5416	0.0000
MV50GP04	-0.0001	0.0000	-8.0393	0.0000
MV50GP05	-0.0017	0.0002	-8.9621	0.0000
MV50GP06	-0.0003	0.0000	-16.8778	0.0000
MV50GP07	-0.0005	0.0002	-2.0503	0.0403
MV50GP08	0.0001	0.0000	8.7203	0.0000
MV50GP09	-0.0001	0.0000	-9.9397	0.0000
MV50GP10	-0.0013	0.0005	-2.7207	0.0065
MV50GP11	0.0001	0.0001	0.5273	0.5980

Residual standard error: 0.782 on 29842 degrees of freedom
Multiple R-Squared: 0.6761
F-statistic: 2831 on 22 and 29842 degrees of freedom, the p-value is 0

MOS	LINEAR REGRESSION MODEL FORMULATION
77F	 

MOS	LINEAR REGRESSION MODEL FORMULATION
88M	**FULL MODEL:** q.88M.Avg.Annl ~ un.rate + MA.POP + EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-6.398	-0.1973	-0.05975	0.1344	19.84

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.1344	0.0135	9.9459	0.0000
un.rate	-1.4912	0.2028	-7.3529	0.0000
MA.POP	0.0001	0.0000	19.8825	0.0000
EXECMNGE	-0.0001	0.0000	-18.1443	0.0000
FAFOFISH	-0.0005	0.0000	-12.6640	0.0000
ADMINSPT	0.0005	0.0000	19.5677	0.0000
PROFSNL	0.0001	0.0000	7.0957	0.0000
TECHSPT	0.0007	0.0000	20.9217	0.0000
SVCOTHR	0.0001	0.0000	6.7995	0.0000
SVCPROT	0.0002	0.0000	4.5486	0.0000
SALES	0.0000	0.0000	2.9086	0.0036
CRFTSMAN	-0.0002	0.0000	-15.6859	0.0000
LABORERS	-0.0017	0.0002	-7.1265	0.0000
TRANSPO	0.0000	0.0000	1.8679	0.0618
MV50GP01	0.0000	0.0000	0.5070	0.6121
MV50GP02	0.0001	0.0000	8.6356	0.0000
MV50GP03	0.0013	0.0000	40.6920	0.0000
MV50GP04	0.0000	0.0000	-8.4558	0.0000
MV50GP05	-0.0016	0.0002	-8.1412	0.0000
MV50GP06	-0.0002	0.0000	-14.0040	0.0000
MV50GP07	-0.0008	0.0002	-3.1417	0.0017
MV50GP08	0.0000	0.0000	7.3058	0.0000
MV50GP09	-0.0001	0.0000	-11.6271	0.0000
MV50GP10	-0.0015	0.0005	-3.1088	0.0019
MV50GP11	0.0001	0.0001	0.5411	0.5884

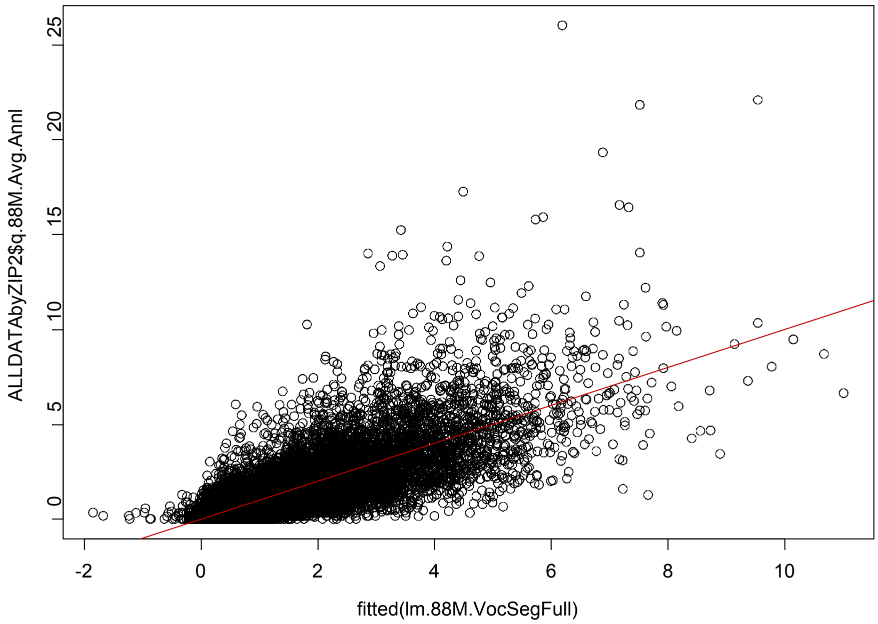
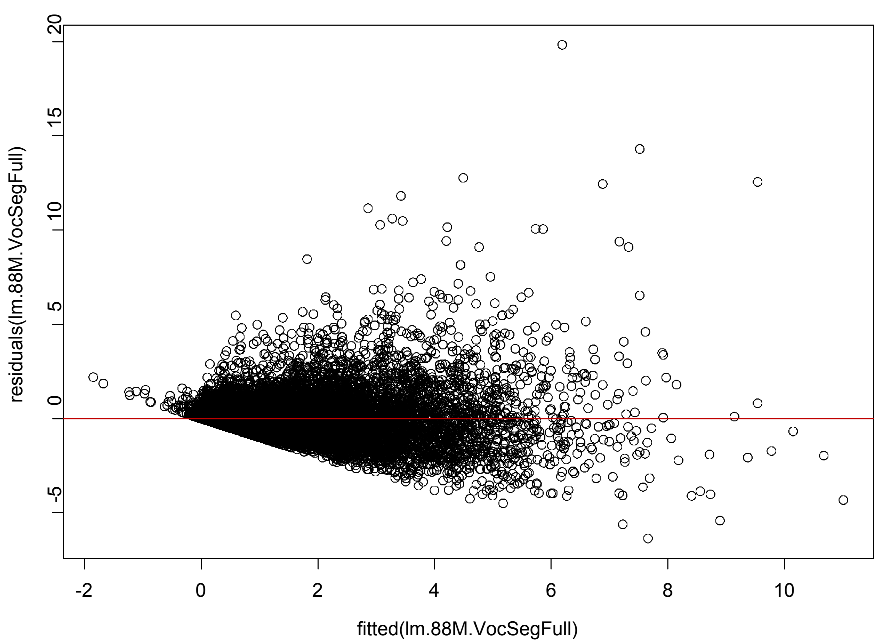
Residual standard error: 0.8012 on 29839 degrees of freedom
Multiple R-Squared: 0.6812
F-statistic: 2656 on 24 and 29839 degrees of freedom, the p-value is 0
1 observations deleted due to missing values

MOS	LINEAR REGRESSION MODEL FORMULATION
88M	**FULL MODEL LESS MA.POP and un.rate:** q.88M.Avg.Annl ~ EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-6.32	-0.1953	-0.0529	0.1273	19.92

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.0461	0.0067	7.0019	0.0000
EXECMNGE	-0.0002	0.0000	-27.4863	0.0000
FAFOFISH	-0.0004	0.0000	-9.8691	0.0000
ADMINSPT	0.0005	0.0000	19.7640	0.0000
PROFSNL	0.0002	0.0000	18.4705	0.0000
TECHSPT	0.0005	0.0000	16.5179	0.0000
SVCOTHR	0.0001	0.0000	13.2621	0.0000
SVCPROT	0.0001	0.0000	2.6673	0.0077
SALES	0.0001	0.0000	8.2037	0.0000
CRFTSMAN	-0.0002	0.0000	-13.6878	0.0000
LABORERS	-0.0018	0.0002	-7.8015	0.0000
TRANSPO	0.0000	0.0000	5.0305	0.0000
MV50GP01	-0.0000	0.0000	-1.7836	0.0745
MV50GP02	0.0000	0.0000	5.4141	0.0000
MV50GP03	0.0013	0.0000	41.0663	0.0000
MV50GP04	-0.0001	0.0000	-7.8656	0.0000
MV50GP05	-0.0017	0.0002	-8.9383	0.0000
MV50GP06	-0.0003	0.0000	-16.7294	0.0000
MV50GP07	-0.0005	0.0003	-2.0636	0.0391
MV50GP08	0.0001	0.0000	9.7702	0.0000
MV50GP09	-0.0001	0.0000	-10.0050	0.0000
MV50GP10	-0.0013	0.0005	-2.7094	0.0067
MV50GP11	0.0001	0.0001	0.5620	0.5741

Residual standard error: 0.807 on 29842 degrees of freedom
Multiple R-Squared: 0.6764
F-statistic: 2835 on 22 and 29842 degrees of freedom, the p-value is 0

MOS	LINEAR REGRESSION MODEL FORMULATION
88M	 

MOS	LINEAR REGRESSION MODEL FORMULATION
95B	**FULL MODEL:** q.95B.Avg.Annl ~ un.rate + MA.POP + EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-5.642	-0.1822	-0.05621	0.1235	17.12

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.1265	0.0123	10.3017	0.0000
un.rate	-1.4385	0.1842	-7.8089	0.0000
MA.POP	0.0001	0.0000	21.9517	0.0000
EXECMNGE	-0.0001	0.0000	-17.8912	0.0000
FAFOFISH	-0.0005	0.0000	-12.4871	0.0000
ADMINSPT	0.0004	0.0000	17.7218	0.0000
PROFSNL	0.0001	0.0000	7.9669	0.0000
TECHSPT	0.0006	0.0000	21.2373	0.0000
SVCOTHR	0.0001	0.0000	5.5187	0.0000
SVCPROT	0.0001	0.0000	3.5368	0.0004
SALES	0.0001	0.0000	4.2069	0.0000
CRFTSMAN	-0.0002	0.0000	-14.9493	0.0000
LABORERS	-0.0015	0.0002	-6.7610	0.0000
TRANSPO	0.0000	0.0000	1.8494	0.0644
MV50GP01	0.0000	0.0000	-0.2632	0.7924
MV50GP02	0.0001	0.0000	9.2359	0.0000
MV50GP03	0.0011	0.0000	37.9182	0.0000
MV50GP04	0.0000	0.0000	-9.4716	0.0000
MV50GP05	-0.0013	0.0002	-7.4895	0.0000
MV50GP06	-0.0002	0.0000	-13.8138	0.0000
MV50GP07	-0.0006	0.0002	-2.6619	0.0078
MV50GP08	0.0000	0.0000	0.7910	0.4290
MV50GP09	-0.0001	0.0000	-12.2948	0.0000
MV50GP10	-0.0014	0.0004	-3.3024	0.0010
MV50GP11	0.0000	0.0001	0.4763	0.6339

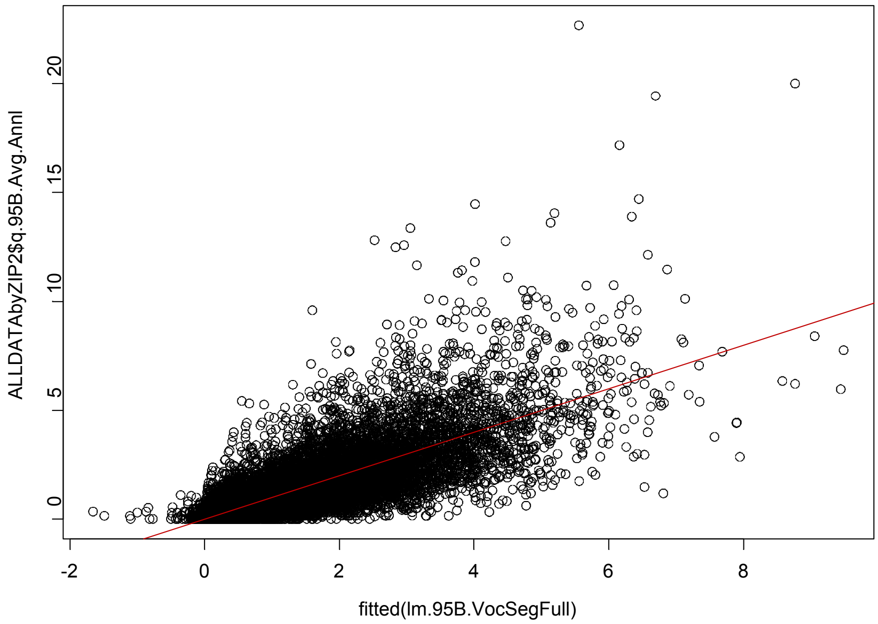
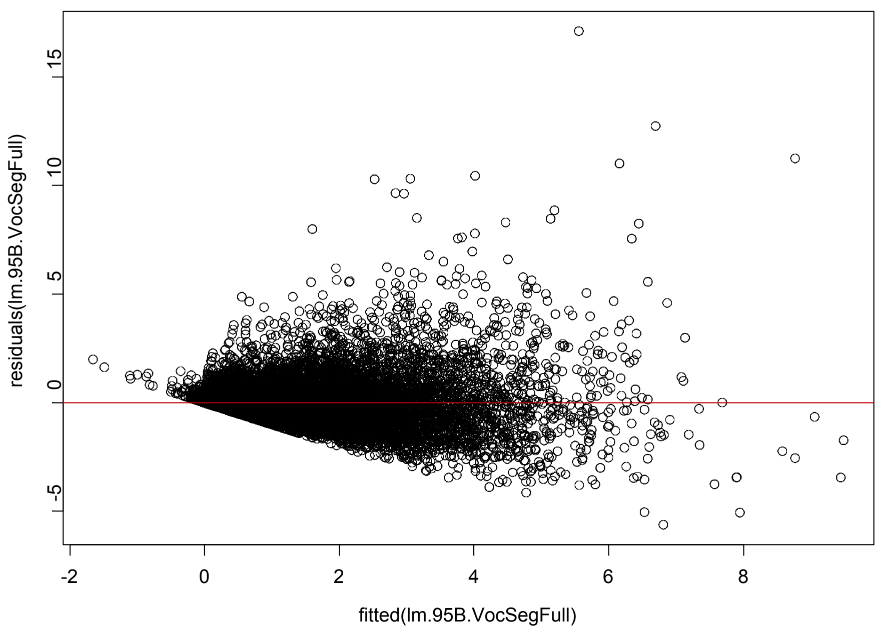
Residual standard error: 0.7278 on 29839 degrees of freedom
Multiple R-Squared: 0.679
F-statistic: 2630 on 24 and 29839 degrees of freedom, the p-value is 0
1 observations deleted due to missing values

MOS	LINEAR REGRESSION MODEL FORMULATION
95B	**FULL MODEL LESS MA.POP and un.rate:** q.95B.Avg.Annl ~ EXECMNGE + FAFOFISH + ADMINSPT + PROFSNL + TECHSPT + SVCOTHR + SVCPROT + SALES + CRFTSMAN + LABORERS + TRANSPO + MV50GP01 + MV50GP02 + MV50GP03 + MV50GP04 + MV50GP05 + MV50GP06 + MV50GP07 + MV50GP08 + MV50GP09 + MV50GP10 + MV50GP11

Residuals:	Min	1Q	Median	3Q	Max
	-5.565	-0.1813	-0.0499	0.1210	17.183

Coefficients	Value	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.0413	0.0060	6.8813	0.0000
EXECMNGE	-0.0002	0.0000	-28.0095	0.0000
FAFOFISH	-0.0004	0.0000	-9.3448	0.0000
ADMINSPT	0.0004	0.0000	17.9164	0.0000
PROFSNL	0.0002	0.0000	20.5149	0.0000
TECHSPT	0.0005	0.0000	16.2856	0.0000
SVCOTHR	0.0001	0.0000	12.5384	0.0000
SVCPROT	0.0001	0.0000	1.4822	0.1383
SALES	0.0002	0.0000	10.0775	0.0000
CRFTSMAN	-0.0002	0.0000	-12.7305	0.0000
LABORERS	-0.0017	0.0002	-7.5061	0.0000
TRANSPO	0.0000	0.0000	5.3468	0.0000
MV50GP01	-0.0000	0.0000	-2.7796	0.0054
MV50GP02	0.0000	0.0000	5.6853	0.0000
MV50GP03	0.0011	0.0002	38.3030	0.0000
MV50GP04	-0.0001	0.0000	-8.7857	0.0000
MV50GP05	-0.0015	0.0002	-8.3361	0.0000
MV50GP06	-0.0002	0.0000	-16.7958	0.0000
MV50GP07	-0.0003	0.0002	-1.4938	0.1352
MV50GP08	0.0000	0.0000	3.4838	0.0005
MV50GP09	-0.0001	0.0000	-10.4622	0.0000
MV50GP10	-0.0012	0.0006	-2.8575	0.0043
MV50GP11	0.0001	0.0001	0.4959	0.6200

Residual standard error: 0.7343 on 29842 degrees of freedom
Multiple R-Squared: 0.6731
F-statistic: 2793 on 22 and 29842 degrees of freedom, the p-value is 0

MOS	LINEAR REGRESSION MODEL FORMULATION
95B	 

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] Bowerman, B. L. and R. T. O'Connell, *Forecasting and Time Series: An Applied Approach,* Third Edition. Belmont, CA: Duxbury Press, 1993.

[2] Gibbons, J. D., *Nonparametric Statistics: An Introduction.* Newbury Park, CA: SAGE Publications, Inc, 1993.

[3] Headquarters, Department of the Army. *Field Manual 5-333: Construction Management*. Washington, DC: Government Printing Office, 1987.

[4] Headquarters, Department of the Army. *Army Regulation 220-1: Unit Status and Reporting*. Washington, DC: Government Printing Office, 1999.

[5] Hogg, R. V. and E. A. Tanis, *Probability and Statistical Inference,* 3rd Edition. New York, NY: Macmillan Publishing Company, 1988.

[6] Hollander, M. and D. Wolfe, *Nonparametric Statistical Methods.* New York, NY: John Wiley & Sons, Inc, 1973.

[7] Scheaffer, R., W. Mendenhall, and L. Ott, *Elementary Survey Sampling,* Fourth Edition. Belmont, CA: Duxbury Press, 1990.

[8] Walpole, R. E. and R. H. Myers, *Probability and Statistics for Engineers and Scientists,* Fourth Edition. New York, NY: Macmillan Publishing Company, 1989.

[9] United States Bureau of Labor and Statistics (Retrieved 28 April 2004). Retrieved from http://www.bls.gov/lau/home.htm#data.

[10] United States Postal Service (Retrieved 14 May 2004). Retrieved from http://www.usps.com/zip4/citytown.htm.

[11] Also used the following analytical tools and software:
Minitab Statistical Software
Microsoft Word
Microsoft Powerpoint
Microsoft Excel
Microsoft Excel Analysis Tools
Microsoft Graph
Microsoft Chart Software
Microsoft FoxPro Relational Database Management System (RDMS)

[12] Litzenberg, Ward (May, 2004). A personal telephonic conversation, RE: USAR Force File, conducted May, 2004.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, VA

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, CA

3. Department of the Army
   Office of the Chief of the Army Reserve
   Washington, DC

4. United States Army Recruiting Command (USAREC)
   ATTN: Program Analysis and Evaluation
   Ft Knox, KY

5. Mr Jamey Christy
   United States Bureau of the Census
   ATTN: Regional Director
   Van Nuyes, CA

6. Dr. David H. Olwell
   Naval Postgraduate School
   Monterey, CA

7. Dr. Samuel E. Buttrey
   Naval Postgraduate School
   Monterey, CA

8. Dr. Charles R. Maderer
   Indiana University of Pennsylvania
   Indiana, PA

9. LTC John Brau
   Naval Postgraduate School
   Monterey, CA

10. MAJ Russ Wyllie
    Naval Postgraduate School
    Monterey, CA

11. Mrs June Grillo
    United States Bureau of the Census
    Los Osos, CA

12. Mrs Susan Fair
    United States Bureau of the Census
    Fredericksburg, VA